■ 特集　小児がん治療の晩期障害と対策

小児がん治療と内分泌合併症

横谷　進*

はじめに

「小児がん」は，白血病や悪性リンパ腫などの血液悪性疾患，脳腫瘍，神経芽腫や横紋筋肉腫などの固形腫瘍を合わせた総称である。小児がんの生命予後は，この40～50年の間に格段に改善し，その80%近くが長期生存できるようになった。これまでは，生命を救うことに最大の関心が払われ，その結果として現在のような高い生存率が達成されてきた。しかし，小児がんそのものに加えて，強力な治療が行われることにより，さまざまな合併症が起こることがわかっている。小児がんを克服した小児がん経験者（childhood cancer survivor：CCS）の一生にとって，そうした合併症は極めて大きな問題として認識されている。北米では，1990年代から大規模な調査が行われ多くの大切な知見が蓄積している[1,2]。

本稿では，CCSにみられる晩期合併症（late effect）のなかで，内分泌合併症に焦点をあて，危険因子に基づく適切なフォローアップと早期診断について述べる。

Ⅰ．晩期合併症としての内分泌合併症の重要性

北米では，若年成人の350～600人に1人がCCSであると言われている[1]。長期生存率がさらに向上してきていることから，今後もしばらくその割合は増加が続くものと推測される。1970～1986年に北米で行われた小児がんに対する治療の内容は，手術療法81.2%，化学療法80.5%，放射線療法67.9%となっており，これらのうち2つ以上の併用療法は85.7%に行われていた。近年は，放射線療法の頻度は低くなってきているが，こうした強力な治療により，内分泌障害，末梢・中枢神経障害，心筋障害，腎障害，運動障害，肝炎，二次がんなど種々の合併症が起こる。これらの晩期合併症のなかでは，内分泌合併症の割合が最も高く，CCSのうちの約40%が内分泌合併症を有すると報告されている[3]。上述のCCSの頻度を考え合わせると，若い成人の約900～1,500人に1人が内分泌合併症を有するCCSであると推測される。種々の内分泌疾患の頻度を考えると，CCSを基盤とするそれが極めて大きな割合を占めることがわかる。CCSにおける内分泌合併症としては，1）成長ホルモン（以下，GH）分泌不全性低身長症と成長障害，2）性腺機能異常，3）その他の下垂体機能不全，4）甲状腺機能低下症，5）骨粗鬆症，6）肥満・メタボリックシンドローム，などが挙げられる。

CCSにとって合併症が実際にどれほどの負担（burden）になっているかを調査した最近の論文によると，中等度以上の負担の大きさを示したイベントの件数は，整形外科学異常，心理社会的・認知能力上の問題に引き続いて，内分泌学的問題が日常生活のなかでも実際に大きな負担になっていることが知られている[4]。

Ⅱ．内分泌合併症を起こす危険因子

CCSにおけるさまざまな合併症は，それを引き起こす危険因子がある場合にのみ，標準リスクより高い確率で起こる。例えば，18 Gy以上の頭部放射線照射が行われていれば，GH分泌不全性低身長症を合併することが十分に予想されるが，通常の化学療法では，その危険は増加しない。すなわち，原病による侵襲と，それ以上に重要である

* 国立成育医療センター第一専門診療部
〔〒157-8535　東京都世田谷区大蔵2-10-1〕

表　主な晩期内分泌合併症とその危険因子

内分泌合併症	放射線療法			化学療法
	頭部	全身	局所	
GH 分泌不全	＋＞18 Gy			
甲状腺機能低下	＋＞40 Gy	＋	＋＞10 Gy	
甲状腺結節/癌	＋	＋	＋＞25 Gy/20-29 Gy	
ACTH 分泌不全	＋＞40 Gy	↓		
思春期早発症	＋＞18 Gy(F)	② 甲状腺と性腺は，放射線感受性が高い		
性腺機能不全・不妊	＋＞40 Gy	＋	＋＞10 Gy（F） ＞1 Gy（M） ＞20 Gy（Leydig）	＋アルキル化剤 ↓ ③ 特定の薬剤が重要
骨粗鬆症	↓ ① 頭部照射が最重要			↑ ＋MTX, ステロイド, BMT
肥満	↑ ＋＞18 Gy			＋ステロイド
高プロラクチン血症	＋＞40 Gy			

MTX：メソトレキセート，BMT：骨髄移植，（主に Nandagopal ら，2008 を参考に作成）[6]

治療内容とによって起こりうる合併症が挙げられ，また，それ以外の合併症は標準リスクを超えて起こる理由がないと判断することができる。こうした考えにより，初めて有効で効率的なフォローアップが可能となり，早期診断・治療に結びつけることができる。このことは，CCS 本人や家族にとっても極めて重要である。無用な心配から解放され，しかし，起こりうる合併症には注意しながら定期的なフォローアップを受けることが期待できる。CCS の長期フォローアップには，このようにリスクに基づいた（risk-based, risk-stratified）対応が不可欠である。

内分泌合併症についても当然，危険因子が知られている。北米の COG（Childhood Oncology Group）では，極めて多数の合併症を挙げて合併症ごとに危険因子や適切なフォローアップについて，情報をウェブ上で公開しているが[5]，そのうちの内分泌合併症については最近のレビューに手際よくまとめられているので参照されたい[6]。

表に，起こりうる主な内分泌合併症と，それぞれに関連する危険因子をまとめた。そこに示したように，概ね 3 つの危険因子に分類することができる。第 1 は，頭部照射である。極めて多くの内分泌合併症（多種の下垂体ホルモン分泌異常）が頭部照射によって引き起される。このことは，脳腫瘍に対してしばしば行われる放射線治療が，内分泌合併症の危険因子として極めて重要であることを示している。しかし，全身照射の一環として行われる頭部照射では，その線量からみて，GH 分泌不全性低身長症以外の視床下部・下垂体機能障害のリスクは上昇しない。第 2 は，甲状腺と性腺に対する放射線照射である。これらの器官は，内分泌器官のなかで放射線感受性が最も高いため，直接の局所照射ではもちろんのこと，全身照射であっても，内分泌合併症は起こる可能性がある。第 3 は，限られた種類の化学療法の薬剤，すなわち，アルキル化剤の一部と，大量・長期のコルチコステロイドである。表からわかるように，放射線合併症が，部位・線量特異的なのに対して，化学療法合併症は，薬剤・用量特異的である。どこにどれだけの放射線照射が行われたか，どの薬剤がどれだけ使われたかは，適切なフォローアップのためには不可欠な情報である。

Ⅲ. 成長障害と GH 分泌不全性低身長症

白血病や脳腫瘍を中心として，CCS では成長障害を生ずることが多い。北米の研究では，小児期の脳腫瘍患者のうちで成人に達した者の平均身長は，同性の同胞に比べて，男子で 10 cm，女子で 7 cm 低かった（それぞれ，$p<0.01$）と報告されている[7]。その原因には，1）GH 分泌不全性低身長症（GHD），2）中枢性または原発性甲状腺機能

低下症，3）中枢性思春期早発症による骨成熟促進，といった内分泌疾患以外に，4）原疾患と集中的治療に伴う低栄養・異化亢進，5）骨（とくに脊椎）の放射線障害などが考えられ，実際にはこれらが組み合わさって成人身長を低くするように作用していることが多い。

GH は，視床下部・下垂体へのさまざまな侵襲に対して最も分泌不全を起こしやすい前葉ホルモンである。わずか 7 Gy の頭部照射でも，小児では GHD が起こりうると言われており，18 Gy を超えれば珍しくなく起こる。40 Gy を超えれば，数年間以内にほぼ 100% に GHD が合併する。このような放射線感受性の高さから，脳腫瘍に対する頭部照射だけでなく，造血幹細胞移植のための全身照射などでも GHD は高い頻度で起こると考えるべきである。放射線照射量が多ければ多いほど短期間に高率で，少ない照射では 10 年にも及ぶ経過で徐々に，GHD が完成する。

小児期の GHD は，成長速度の低下で見つけることができる。原疾患や治療によりその可能性があると判断される CCS では，成長曲線を描きながらフォローし，標準成長曲線に比べて成長速度（成長曲線の傾き）が低下してきた場合には，GHD を疑うべきである。GHD の診断には，身長によらず成長速度が 2 年以上にわたって −1.5 SD 以下であれば成長障害の要件を満たす。脳内の器質的原因や頭部放射線照射の既往のある CCS では，GH 分泌刺激試験は 1 種類で十分であると考えられている（厚生労働省研究班による診断の手引き[8]，および，小児慢性特定疾患の認定基準）。成長障害がありながら，GH 分泌低下が認められない場合にも，GHD は年の単位で完成するので，しばらく後に再検査を行うと分泌低下を示すことも珍しくない。こうしたことを熟知して，適切な症例には，遅れることなく GH 治療が積極的に考慮されるべきである。

IV. ゴナドトロピン-性腺系の機能異常

ゴナドトロピン（Gn）-性腺系の異常は，次項・次々項で述べられるので簡単に述べる。頭部照射では，18 Gy を超えると女子において思春期早発症をきたすことが珍しくない。30 Gy を超えると男子でもその可能性がある。40 Gy を超えると，逆に GnRH 分泌が障害されて，しばしば中枢性性腺機能不全を起こす。また，40 Gy 以上では，視床下部による抑制が解除されて高プロラクチン血症をきたすことがあり，これも性腺機能を抑制する原因になりうる。

一方，性腺自体への直接の障害では，アルキル化剤（とくに，busulfan と cyclophosphamide）の使用が危険因子として知られている。放射線療法では，わずか 1 Gy でも精子形成障害が起こり，4 Gy を超えると永続的な無精子症となることが多い。Leydig 細胞障害（すなわちテストステロン分泌障害）は，20 Gy 以上が危険因子である。テストステロン分泌障害では，思春期の遅発や停止として現われるので，危険因子がある場合には二次性徴に注意してフォローする必要がある。卵巣の放射線感受性は思春期開始前と後では異なっていて，思春期前では 10〜15 Gy を超えると危険因子となるが，思春期開始後では 5〜10 Gy 以上で卵巣機能不全が起こりうる。アルキル化剤に対する感受性も同様で，思春期開始後のほうが障害されやすい。

男女いずれの場合も，性腺機能不全では，中枢性にしても原発性にしても，ホルモン補充療法が適切に行われることが必要である。それは，二次性徴や性機能のためだけではなく，骨密度の増加・維持のために不可欠と考えられるからである。小児期には，二次性徴の完成までを目指す段階的な治療により補充療法が開始される[9]。

V. その他の内分泌合併症

下垂体機能では，40 Gy（とくに 50 Gy）を超えた頭部照射で，GH に加えて TSH，ACTH，時には，ADH の分泌不全を合併することが少なくない。

甲状腺自体も放射線の影響を受けやすい。10 Gy を超えると，甲状腺機能低下症，時に亢進症を発症する。25 Gy を超えると甲状腺結節を起こしやすい。甲状腺癌は 20〜29 Gy で最も起こりやすいと報告されている。

骨粗鬆症は，アルキル化剤（とくに，メソトレキセート）とコルチコステロイド（とくに，プレ

ドニゾンとデキサメサゾン）が危険因子となって起こりやすい。とくに，これらの併用療法，GVHDなどでのコルチコステロイドの長期投与，GHD・性腺機能低下・甲状腺機能亢進症などの内分泌合併症，喫煙・アルコール・運動不足・カルシウム摂取不足などの生活習慣が危険を高める。

VI. CCSへの最近の取り組み

国内では，日本小児白血病・リンパ腫治療研究グループ（JPLSG）の長期フォローアップ委員会（石田也寸志委員長）と，厚生労働省の研究班（石田班）によって，身体的心理的晩期合併症やQOLの後方視的な調査が行われ，フォローアップガイドの翻訳，CCSのための健康手帳や共通治療サマリーの制作などが進行している。内分泌合併症については，日本小児内分泌学会のなかに，CCSの内分泌合併症に関する委員会（委員長：横谷　進）が2006年9月に発足し，長期フォローアップ委員会・研究班と歩調を合わせて活動し，現在は，内分泌合併症の早期診断と介入のためのガイドラインを作成中である。さらに，2007年には，コンセンサス作成・小児がんフォローアップ拠点病院モデル事業・小児がん登録の基盤整備を3本柱とする，新たな研究班（主任研究者：藤本純一郎）が発足した。このように，国内ではようやく本格的な取り組みが始まったところである。

VII. 2つのトランジション

小児がんの治療は，小児血液・腫瘍科医・小児外科医・脳外科医など，腫瘍の専門家により行われ，通常は，これらの人々が中心となってフォローが行われている。起こりうる合併症に焦点をあてた適切なフォローアップの間に合併症（の疑い）が生じれば，専門家に紹介されるが，これは，第1の移行（トランジション）と呼ばれる。この移行がスムーズに行われるためには，適切なフォローアップとリファーの基準，そして，何よりもフォローアップ医と合併症専門医（小児内分泌科医）の日常的な協力関係が重要である。第2のトランジションは，小児期から成人期への移行である。内分泌合併症は，その多くが永続的であることから，いずれは成人内分泌科医による診療に移行しなければならない。GH分泌不全について，日本小児内分泌学会によるトランジションのガイドラインがある[10]。成人診療医にも，CCSの合併症と生涯にわたるケアについての知識と理解が求められる[11]。

おわりに

本稿の主旨は以下のとおりである。

1）小児がん経験者（CCS）のなかで，内分泌合併症の頻度は高い。

2）種々の内分泌合併症の可能性は，原病と治療内容により合併症ごとに予測できる。内分泌合併症の危険因子のなかでは，頭部放射線照射，甲状腺・性腺への照射，一部のアルキル化剤と長期・大量のコルチコステロイドが重要である。

3）リスクに基づいたフォローアップと，小児内分泌科医への適切なリファーが強く望まれる。

4）成人診療医へのトランジションが準備されなければならない。

文　献

1) 石田也寸志：北米 Childhood Cancer Survivors Studyによる小児がん経験者の長期的な問題点，第1編．日小児会誌 **110**：1513-1522，2006
2) 石田也寸志：北米 Childhood Cancer Survivors Studyによる小児がん経験者の長期的な問題点，第2編．日小児会誌 **110**：1523-1533，2006
3) Stevens MC, Mahler H, Parkes S：The health status of adult survivors of cancer in childhood. Eur J Cancer **34**：694-698, 1998
4) Geenen MM, Cardous-Ubbink MK, Kremer LCM, et al：Medical assessment of adverse health outcomes in long-term survivors of child cancer. JAMA **297**：2705-2715, 2007
5) http://www.childrensoncologygroup.org/
6) Nandagopal R, Laverdière C, Mulrooney D, et al：Endocrine late effects of childhood cancer therapy：a report from the Children's Oncology Group. Horm Res **69**：65-74, 2008
7) Gurney JG, Ness KK, Stovall M, et al：Final height and body mass index among adult survivors of childhood brain cancer：childhood cancer survivor study. J Clin Endocrinol Metab **88**：4731-4739, 2003
8) 厚生労働省難治性疾患克服研究事業「間脳下垂体機能障害に関する調査研究」：成長ホルモン分泌不全

性低身長症の診断の手引き（平成19年度改訂）
9) 横谷　進：小児思春期における補充療法．日内分泌会誌 82（増刊）：53-55, 2006
10) 横谷　進，依藤　亨，田中敏章，他，日本小児内分泌学会成長ホルモン委員会：成長ホルモン分泌不全性低身長症の小児期の成長ホルモン治療から成人期の成長ホルモン治療への移行ガイドライン．日小児会誌 110（10）：1475-1479, 2006
11) 横谷　進：第18回臨床内分泌代謝 Update「臨床医に必要な Childhood Cancer Survivors（CCS）の知識」．日内分泌会誌 84（増刊）：掲載予定，2008

Childhood Cancer Therapy and Endocrine Complications

SUSUMU YOKOYA

Department of Medical Subspecialties, National Center for Child Health and Development

Key words：Childhood cancer survivors, Late effects, Endocrine complications, Risk-based follow-up, Transition.
Jpn. J. Pediatr. Surg., 40(6)；671～675, 2008.

The incidence of the endocrine complications is the most common of the late effects observed in childhood cancer survivors (CCS). Occurrence of various endocrinopathies can be predicated by their specific risk factors, of which the most important are cranial irradiation, radiation exposure to thyroid or sex glands, and chemotherapy including use of alkylating agents and corticosteroids. Thus risk-based follow-up and adequate referral to pediatric endocrinologists (first transition) are essential. This will assure CCSs' growth, sexual development, and acquisition of bone mineral content to remain normal. Transition from pediatric to adult care (second transition) is another issue to be resolved.

*　　*　　*

Association of a missense single nucleotide polymorphism, Cys1367Arg of the WRN gene, with the risk of bone and soft tissue sarcomas in Japan

Robert Nakayama,[1,2] Yasunori Sato,[1] Mitsuko Masutani,[3] Hideki Ogino,[3] Fumihiko Nakatani,[4] Hirokazu Chuman,[4] Yasuo Beppu,[4] Hideo Morioka,[2] Hiroo Yabe,[2] Hiroshi Hirose,[5] Haruhiko Sugimura,[6] Hiromi Sakamoto,[1] Tsutomu Ohta,[1] Yoshiaki Toyama,[2] Teruhiko Yoshida[1,7] and Akira Kawai[4]

[1]Genetics Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045; [2]Department of Orthopaedic Surgery, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582; [3]ADP-ribosylation in Oncology Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045; [4]Division of Orthopedic Surgery, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045; [5]Health Center, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582; [6]Department of Pathology, Hamamatsu University School of Medicine, 1-20-1 Handayama, Hamamatsu-shi, Shizuoka, 431-3192 Japan

(Received June 5, 2007/Revised October 3, 2007/Accepted October 15, 2007/Online publication February 4, 2008)

Bone and soft tissue sarcomas (BSTSs) are rare malignant tumors of mesenchymal origin. Although BSTSs frequently occur in some hereditary cancer syndromes with germline mutations of DNA repair genes, genetic factors responsible for sporadic cases have not been determined. In the present study we undertook a case-control study and analyzed possible associations between the susceptibility to BSTS and the single nucleotide polymorphisms (SNPs) in DNA repair genes. Genomic DNAs extracted from case and control peripheral blood leukocytes were genotyped by pyrosequencing. For candidate polymorphisms, we chose 50 non-synonymous missense SNPs, which we have previously been identified by resequencing 36 DNA repair genes among the Japanese population. In the first screening, we analyzed 240 cases and 685 controls and selected six SNPs at the significance level of $P < 0.1$ (Fisher's exact test). The six SNPs were further analyzed in the second genotyping on an additional set of 304 cases and 834 controls. In the joint analysis (the first and second genotyping combined) of 544 cases and 1378 controls, Cys1367Arg of the WRN gene was found to be a protective factor of BSTS (odds ratio = 0.66, 95% confidence interval = 0.49–0.88, $P = 0.005$). An exploratory subgroup analysis without multiple comparison adjustment suggested that the WRN-Cys1367Arg SNP is associated with soft tissue sarcomas, sarcomas with reciprocal chromosomal translocations and malignant fibrous histiocytoma. (*Cancer Sci* 2008; 99: 333–339)

Bone and soft tissue sarcomas (BSTSs) are nonepithelial, non-hematological malignant tumors of mesenchymal origin. Three characteristics of BSTS are that: (i) they are rare malignancies; (ii) they develop at a variety of sites; any mesenchymal tissues throughout the body; and (iii) show highly heterogeneous histological types. In the United States, the incidence of malignant bone tumors is estimated to be around 0.8 per 100 000 population,[1] and that of soft tissue sarcomas (STS) approximately 5.0 per 100 000.[2] It is estimated that 2370 bone and joint malignancies and 9220 malignant soft tissue tumors would newly develop in 2007.[3]

The rarity and the histological heterogeneity of BSTSs have hampered the identification of the risk factors and etiology. BSTSs show a slight male predominance, and the incidence of these rare tumors both in Japan and in other countries seems to be stable in the last several decades except for an increase in Kaposi sarcoma as reported in the United States.[4,5] This appears to be in contrast to several other types of cancers, such as gastrointestinal or gynecologic cancers, which have shown a significant change in incidence in Japan probably due to the change in environmental and life style factors. No significant racial variation has been noted in the overall incidence of sarcomas with some exceptions, such as Ewing sarcoma, which reportedly occurs more frequently in Caucasians.[6]

To date, some environmental and genetic factors for BSTS risk have been suggested. The environmental factors include external radiation therapy,[7] Thorotrast, arsenical pesticides and medications, phenoxyherbicides, dioxin, vinyl chloride, immunosuppressive drugs, alkylating agents, androgen-anabolic steroids, human immunodeficiency virus, and human herpes virus type 8.[5] Information on the genetic factors has so far been limited to certain monogenic hereditary cancer syndromes known to be associated with the incidence of BSTS, such as Li–Fraumeni syndrome, hereditary retinoblastoma, and Werner syndrome with a germline mutation of *TP53*, *RB1*, and *WRN*, respectively. These genes are involved in DNA repair and related systems. Other monogenic hereditary syndromes are also associated with the specific type of multiple benign tumors and their malignant transformation, such as multiple neurofibromas and malignant peripheral nerve sheath tumors in neurofibromatosis 1, and multiple osteochondromas and chondrosarcomas in hereditary multiple exostosis, that have a germline mutation of *NF1* and *EXT1/2*, respectively.

Recently, BSTSs have been considered to be divided into two distinct entities based on their somatic genetic aberrations. One group is characterized by reciprocal chromosome translocations resulting in tumor-specific fusion genes, which may be a critical step for pathogenesis. Another group of sarcomas tend to show complex abnormalities in the karyotypes, suggesting an overall increase in genetic and chromosomal instability. However, the precise mechanisms of tumorigenesis in both groups remain unclear.

The characteristic increase in the risk of BSTS in inherited diseases caused by germline mutations in the DNA repair and related systems has prompted us to investigate the association of the DNA repair gene polymorphisms with the risk of BSTSs. A case-control study was carried out on 544 cases with BSTS and 1378 controls at 50 non-synonymous coding single nucleotide polymorphisms (cSNPs), which we have identified by resequencing of 36 candidate genes on a Japanese population[8]. Associated studies of common polymorphisms of the DNA repair genes have already been reported in many types of cancer,[9] but to our knowledge, this study is the first report on BSTS.

[7]To whom correspondence should be addressed. E-mail: tyoshida@ncc.go.jp

doi: 10.1111/j.1349-7006.2007.00692.x

Table 1. Fifty missense single nucleotide polymorphisms (SNPs) in DNA repair genes grouped by their representative pathways

DNA repair gene	Gene name	SNP	Amino acid change
Base excision repair			
PARP-1/ADPRT1	Poly (ADP-ribose) polymerase family, member 1	T2444C	Val762Ala
		A2978G	Lys940Arg
APE1/APEX	APEX nuclease (multifunctional DNA repair enzyme) 1	A395G	Ile64Val
		T649G	Asp148Glu
MBD4	Methyl-CpG binding domain protein 4	G1212A	Glu346Lys
NUDT1	Nudix (nucleoside diphosphate linked moiety X)-type motif 1	G273A	Val83Met
OGG1	8-oxoguanine DNA glycosylase	C2243G	Ser326Cys
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1	C685T	Arg194Trp
		G944A	Arg280His
		G1301A	Arg399Gln
Nucleotide excision repair			
ERCC5/XPG	Excision repair cross-complementing rodent repair deficiency, complementation group 5	C3507G	His1104Asp
ERCC6/CSB	Excision repair cross-complementing rodent repair deficiency, complementation group 6	G1275A	Gly399Asp
XPC	Xeroderma pigmentosum, complementation group C	A2655C	Lys822Gln
XPD/ERCC2	Excision repair cross-complementing rodent repair deficiency, complementation group 2	G1615A	Asp312Asn
		A2932C	Lys751Gln
Mismatch repair			
MLH1	mutL homolog 1, colon cancer, non-polyposis type 2	A676G	Ile219Val
MLH3	mutL homolog 3 (*Escherichia coli*)	C2645T	Pro844Leu
		C2939T	Thr942Ile
MSH2	mutS homolog 2 (E. coli)	C91T	Thr8Met
MSH3	mutS homolog 3 (E. coli)	A3122G	Thr1036Ala
MSH6	mutS homolog 6 (E. coli)	G203A	Gly39Glu
DNA damage response genes			
TP53	Tumor protein p53 (Li–Fraumeni syndrome)	G466C	Arg72Pro
DNA double strand break repair			
BLM	Bloom syndrome	C967T	Thr298Met
		G4035A	Val1321Ile
BRCA2	Breast cancer 2, early onset	A1342C	Asn372His
KIAA0086	DNA cross-link repair 1A	C1867G	His317Asp
LIG4	DNA ligase IV	A2245G	Ile591Val
NBS1	Nijmegen breakage syndrome 1	G605C	Glu185Gln
RAD51L3	RAD51-like 3 (S. cerevisiae)	G501A	Arg126Gln
RAD54L	RAD54-like (S. cerevisiae)	A551G	Lys151Glu
RINT1	RAD50 interactor 1	G33C	Glu4Gln
WRN	Werner syndrome	C2573T	Thr781Ile
		T4330C	Cys1367Arg
XRCC3	X-ray repair complementing defective repair in Chinese hamster cells 3	C1075T	Thr241Met
DNA polymerase			
POLD1	Polymerase (DNA directed), delta 1	G409A	Arg119His
POLH/XPV/RAD30	Polymerase (DNA directed), eta	A1840G	Lys535Glu
POLI/RAD30B	Polymerase (DNA directed) iota	A2180G	Thr706Ala
POLL	Polymerase (DNA directed), lambda	C1683T	Arg438Trp
POLZ/REV3	REV3-like, catalytic subunit of DNA polymerase zeta (yeast)	C4259T	Thr1146Ile
REV1	REV1 homolog (S. cerevisiae)	T982C	Phe257Ser
		A1330G	Asn373Ser
Other pathways			
FANCA	Fanconi anemia, complementation group A	G827A	Ala266Thr
		G1080A	Arg350Gln
		A1532G	Ser501Gly
		A2457G	Asp809Gly
		C3294T	Ser1088Phe
FANCE	Fanconi anemia, complementation group E	G451T	Arg89Leu
		G1213A	Arg343Gln
FANCF	Fanconi anemia, complementation group F	A983G	Lys324Glu
FANCG/XRCC9	Fanconi anemia, complementation group G	C1382T	Thr297Ile

ADP, adenosine diphosphate;

doi: 10.1111/j.1349-7006.2007.00692.x

Table 2. Case and control subjects

		1st screening		2nd genotyping		Joint analysis	
		Case	Control	Case	Control	Case	Control
Gender	Male	143	483	174	492	317	875
	Female	97	202	130	342	227	503
Age	<40	109	113	127	214	236	318
	≥40	131	572	177	620	308	1060
Institutions[1]	NCCH	240	242	127		367	242
	NCCE			19		19	
	KEIO		302	158	507	158	809
	IWT				327		327
	NNH		141				
Total		240	685[1]	304	834	544	1378[2]

Distributions of gender, age and five institutions where case and/or control subjects were recruited are listed. [1]The 685 controls in the first screening are from our published data (8). [2]The 141 NNH control data were not available for the joint analysis. KEIO, Keio University Hospital, Tokyo; IWT, Iwata Hospital, Shizuoka; NCCE, National Cancer Center East, Chiba; NCCH, National Cancer Center Hospital, Tokyo; NNH, National Nishigunma Hospital, Gunma.

Table 3. Histological distribution of the cases and subgroup classification

Bone sarcoma	
Osteosarcoma	105
Chondrosarcoma	41
EWS/PNET	18
MFH	11
Chordoma	8
Adamantinoma	3
Leiomyosarcoma	2
MPNST	1
Fibrosarcoma	1
	190

Soft tissue sarcoma	
Liposarcoma	111
MFH	92
Synovial sarcoma	38
Leiomyosarcoma	19
MPNST	19
Dermatofibrosarcoma protuberance	17
EWS/PNET	15
EMC	7
Osteosarcoma	6
Fibrosarcoma	5
Epithelioid sarcoma	5
Rhabdomyosarcoma	6
Angiosarcoma	6
Alveolar soft part sarcoma	4
Clear cell sarcoma	2
Mesenchymal chondrosarcoma	1
Hemangiopericytoma	1
	354

Subgroup	Total (%)
ALL	544 (100.0)
BS/STS	
BS	190 (34.9)
STS	354 (65.1)
Translocation	
(+)	154 (28.3)
(–)	390 (71.7)
Histopathology	
Osteosarcoma	111 (20.4)
MFH	103 (18.9)
Liposarcoma	111 (20.4)
Others	219 (40.3)

ALL, all samples; BS, bone sarcomas; EMC, extraskeletal myxoid chondrosarcoma; EWS/PNET, Ewing sarcoma and peripheral neuroectodermal tumor; MFH, malignant fibrous histiocytoma; MPNST, malignant peripheral nerve sheath tumor; STS, soft tissue sarcomas; Translocation (+), sarcomas with reciprocal chromosomal translocations; Translocation (–), sarcomas without reciprocal chromosomal translocations.

Materials and Methods

Study design, case and control subjects. Genotype data of the 50 SNPs,[8] in Table 1 from 240 cases and 685 controls were analyzed as the first screening to select those SNPs to be subjected to the second genotyping in the additional 304 cases and 834 controls. In this study, we used a joint analysis,[10] in which the results of the second genotyping were combined with those of the first genotyping (Table 2).

The cases were recruited in three hospitals (National Cancer Center Hospital [NCCH], Tokyo, National Cancer Center East Hospital, Chiba [NCCE], and Keio University Hospital, Tokyo [KEIO]) from January 2004 to March 2005 (Table 2). The patients, all Japanese, were either newly diagnosed as BSTS or had been followed up in an outpatient clinic for the history of BSTS. The cases consisted of various histological subtypes, which were confirmed by histological examination in each hospital (Table 3).

We used 685 control subjects in the first screening, which were the same control population as those used in our previous study on the same set of SNPs but on a different type of malignancy, lung cancer (Table 2).[8] Those subjects consisted of 383 non-cancer patients in two hospitals (NCCH and National Nishigunma Hospital, Gunma [NNH] in Table 2) and 302 healthy volunteers in KEIO (Table 2). We did not, however, include the 141 control subjects from NNH (Table 2) in the subsequent joint analysis, because their individual genotyping data have not been published and are unavailable for analyses other than the lung cancer study. In the second genotyping, we analyzed additional samples from 834 people who participated in the health examination programs (KEIO and Iwata Hospital, Shizuoka [IWT] in Table 2). We consider that these subjects were suitable enough as controls for our cases; most of the control subjects in this study were also analyzed in our separate gastric cancer project involving a SNP-based genome scan, and its data suggested little if any population stratification (unpublished data, 2007). Therefore, the control subjects for the joint analysis totaled 1378 people with a criterion of no history of cancer during the study period (Table 2).

Subgroup analysis was carried out on the subgroups defined by Table 3. For the 'sarcomas with reciprocal chromosomal

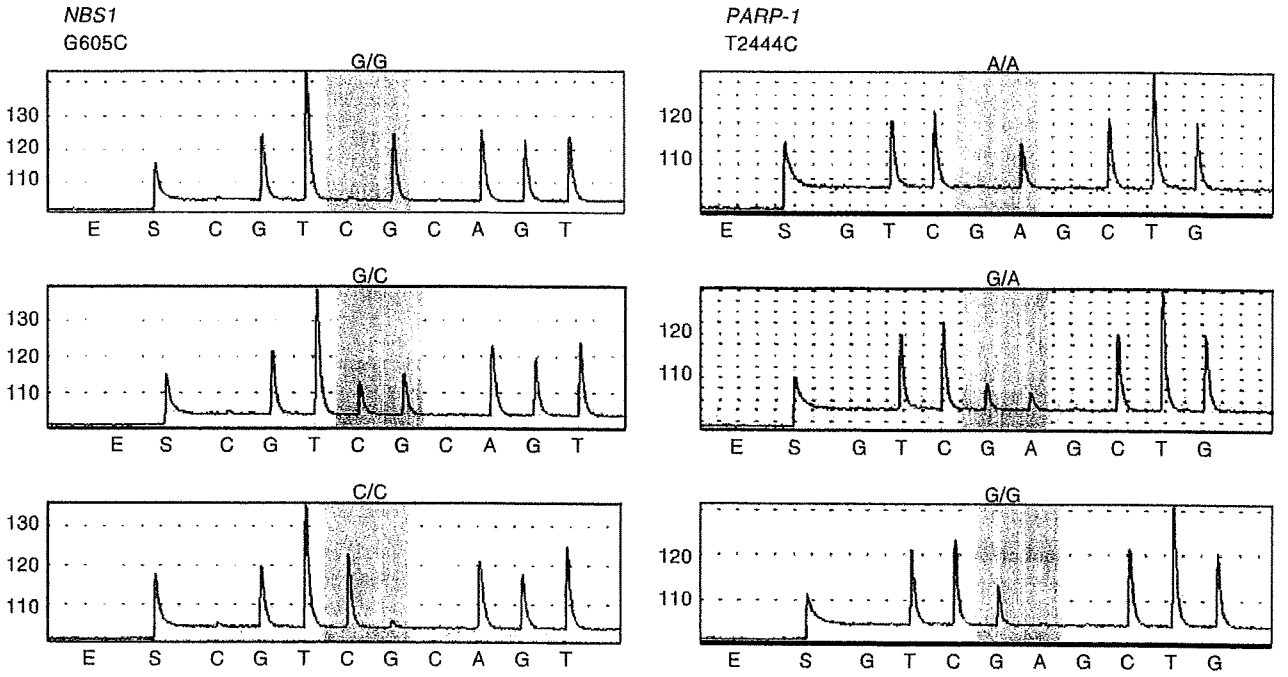

Fig. 1. Genotyping by pyrosequencing. (a) NBS1-Glu185Gln (G605C) and (b) PARP-1-Val762Ala (T2444C) (sequencing in reverse direction). Top, major homozygote; middle, heterozygote; and bottom, minor homozygote.

translocations', we included synovial sarcoma, Ewing sarcoma/peripheral neuroectodermal tumor, myxoid/round cell liposarcoma, clear cell sarcoma, dermatofibrosarcoma protuberance, extraskeletal myxoid chondrosarcoma, and alveolar soft part sarcoma. The specific fusion genes have been reported for those sarcomas, and cytogenetic examinations are routine for their histological diagnosis.[11]

This study was approved by Institutional Review Board of each institution, and all of the subjects signed an informed consent form to participate in the study.

DNA extraction and genotyping. From each individual we obtained a 10–20-mL sample of whole blood. Genomic DNAs were isolated directly from the samples using Blood Maxi Kit (Qiagen, Tokyo, Japan) or FlexiGene DNA Kit (Qiagen) according to the manufacturer's instructions. Ten nanograms of genomic DNA were subjected to genotyping for 50 SNPs by pyrosequencing using the PSQ96 system (Pyrosequencing, Uppsala, Sweden) as described previously.[12] Briefly, a genomic fragment containing an SNP site was amplified by polymerase chain reaction (PCR) with a set of PCR primers, one of which was biotinylated. The PCR products were purified using streptavidin-modified paramagnetic beads (Dynabeads M-280; Dynal, Skoyen, Norway), denatured and subjected to nucleotide sequencing by pyrosequencing chemistry. Quality of the SNP typing was confirmed by inspection of the sequence data and by Hardy–Weinberg equilibrium (HWE) tests.

Statistical analysis. Fisher's exact test, odds ratios (OR) and 95% confidence intervals (CI) were used to analyze association of the SNPs and BSTS risk in allele, dominant (i.e. aa + aA vs AA, where 'A' is major allele and 'a' is minor allele) and recessive (i.e. aa vs aA + AA) models.[13] Crude OR was used for the allele model, while the dominant and recessive models were analyzed using OR adjusted for age (≥40 vs <40, see Suppl. Fig. S1 online) and gender with 95% CI calculated using a logistic regression analysis. When statistical calculation is not applicable due to 0 subjects being in a cell of a contingency table, we indicate the result as 'NA'.

In order to identify disease-associated SNPs, the following criteria were used: candidate SNPs were selected by P-value < 0.10 on the allele model in the first screening, and then disease associated-SNPs were statistically identified by the allele model in the joint analysis with a multiple comparison adjustment using the Holm's method,[14] for six SNPs. Please note that the family wise error rate may be inflated by a few percent, depending on the number of the true SNPs present in the initial 50 SNPs, by this method of multiple comparison adjustment, because the first and second screenings are not independent in the joint analysis (see Suppl. Table S1 online).

A subgroup analysis was carried out as an exploratory, adjunct analysis without multiple comparison adjustment to address the histological heterogeneity of BSTS. Because the subgroups showed different age preferences, dominant and recessive models were used in order to adjust age and gender by multiple logistic regression.

Results

Results of the joint analysis. The typical pyrosequencing data are shown in Fig. 1, and the results of the two-stage genotyping are summarized in Table 4. Minor allele frequencies and P-values of the HWE tests are also listed in Table 4. The six SNPs were selected by the first screening, and the final statistical gene selection was made using all the samples genotyped in the first screening and the second genotyping combined, except 141 controls from Nishigunma Hospital (cases = 544 and controls = 1378 in total). We identified a SNP, WRN-Cys1367Arg, whose allele frequency was significantly different between all BSTS cases and the control subjects (OR = 0.66, 95% CI = 0.49–0.88, $P = 0.005$ and $P = 0.03$, before and after six-SNP multiple comparison adjustment by Holm's method, respectively), showing a protective effect. The minor allele frequency of the SNP is 8.2%.

Subgroup analysis. Since BSTS is characterized by the substantial heterogeneity in its histology, the effects of the DNA repair gene polymorphisms might differ among the subgroups. The results of the subgroup analysis of the WRN-Cys1367Arg SNP are shown in Table 5. Although under-powered and exploratory in nature, the subgroup analysis suggested that the difference in genotype frequency of WRN-Cys1367Arg between the cases and controls

doi: 10.1111/j.1349-7006.2007.00692.x

Table 4. Statistics of allele model analysis for the single nucleotide polymorphisms (SNPs) selected in the first screening by $P < 0.1$

Gene	SNP		Minor allele frequency		HWE (*P*-value)		OR	Allele model	
			Cases (*n*)	Controls (*n*)	Cases	Controls		95% CI	*P*-value
MBD4	G1212A	1st screening	0.304 (240)	0.349 (685)	0.545	0.207	0.82	0.65–1.02	0.083
	Glu346Lys	Joint analysis	0.335 (544)	0.353 (1378)	0.620	0.199	0.92	0.79–1.07	0.286
MSH6	G203A	1st screening	0.275 (240)	0.323 (685)	0.074	0.793	0.80	0.63–1.00	0.057
	Gly39Glu	Joint analysis	0.304 (544)	0.310 (1378)	0.434	0.621	0.97	0.84–1.13	0.750
PARP-1	A2978G	1st screening	0.077 (240)	0.050 (685)	1.000	1.000	1.58	1.04–2.38	0.044
	Lys940Arg	Joint analysis	0.066 (544)	0.052 (1378)	1.000	0.509	1.30	0.97–1.74	0.098
REV1	A1330G	1st screening	0.023 (240)	0.043 (685)	1.000	1.000	0.52	0.27–1.00	0.055
	Asn373Ser	Joint analysis	0.029 (544)	0.044 (1378)	0.042	0.680	0.67	0.45–0.99	0.049
WRN	T4330C	1st screening	0.056 (240)	0.090 (685)	1.000	0.480	0.61	0.40–0.93	0.024
	Cys1367Arg	Joint analysis	0.052 (544)	0.082 (1378)	0.213	0.619	0.66	0.49–0.88	0.005
XRCC1	C685T	1st screening	0.279 (240)	0.327 (685)	0.016	1.000	0.80	0.63–1.00	0.058
	Arg194Trp	Joint analysis	0.292 (544)	0.323 (1378)	0.587	0.811	0.87	0.74–1.01	0.073

CI, confidence interval; HWE, Hardy–Weinberg equilibrium; OR, odds ratio.

Table 5. Statistics of subgroup analysis using recessive and dominant models

SNP	Subgroup	(*n*)	Recessive model†			Dominant model‡		
			OR	95% CI	*P*-value	OR	95% CI	*P*-value
WRN	ALL	544	0.43	0.09–2.01	0.442	0.66	0.48–0.91	0.011
T4330C	BS	190	NA	NA–NA	0.489	0.85	0.53–1.37	0.589
Cys1367Arg	STS	354	0.72	0.16–3.32	0.998	0.58	0.40–0.85	0.005
	Translocation (+)	154	NA	NA–NA	0.632	0.52	0.29–0.95	0.034
	Translocation (–)	390	0.61	0.13–2.86	0.817	0.72	0.51–1.02	0.068
	Osteosarcoma	111	NA	NA–NA	0.826	1.00	0.56–1.80	1.000
	MFH	103	NA	NA–NA	1.000	0.45	0.22–0.94	0.032
	Liposarcoma	111	NA	NA–NA	0.911	0.72	0.40–1.31	0.346

Odds ratios (OR) adjusted for age (≥ 40, < 40) and gender with 95% confidence intervals (CI) were calculated using a logistic regression analysis for all subgroups. Cases = 544 and Controls = 1378 in total. †Recessive model is aa versus aA + AA. ‡Dominant model is aa + aA *versus* AA, where a is a minor allele. ALL, all samples; BS, bone sarcomas; MFH, malignant fibrous histiocytoma; NA, statistical calculation not applicable, because the number of the subjects is less than five in any cell in a contingency table; SNP, single nucleotide polymorphism; STS, soft tissue sarcomas; Translocation (+), sarcomas with reciprocal chromosomal translocations; Translocation (–), sarcomas without reciprocal chromosomal translocations.

appears to be significant in STS, sarcomas with reciprocal chromosomal translocations and malignant fibrous histiocytoma. It is noted, however, that the effect direction of the minor allele (Arg1367) of this SNP was protective in all of the subgroups, except osteosarcoma (OR = 1.00), irrespective of the statistical significance. The subgroup analysis data of the other five SNPs analyzed in the second stage is shown in Suppl. Table S2 online.

Discussion

This study attempted the first systematic survey on the possible role of DNA repair gene polymorphisms in the susceptibility to sporadic BSTS. From the joint analysis of the two-stage case-control study on total 544 cases with BSTS of various histology and 1378 controls, a missense SNP of the WRN gene, Cys1367Arg, was identified. WRN is a member of the RecQ family of DNA helicases, and mutations of the gene can give rise to a rare autosomal recessive genetic instability disorders, Werner syndrome (WS).[15] WS is a premature aging disease characterized by predisposition to cancer and the early onset of symptoms related to normal aging.[16] The types of cancer with elevated risk appear selective, including soft tissue sarcomas, thyroid carcinoma, malignant melanoma, meningioma, hematological malignancies, and osteosarcoma. A diversity of malignancies was found in WS in the literature from 1939 to 1995, but it is noteworthy that the ratio of epithelial to nonepithelial cancers was about 1:1, instead of the usual 10:1.[17] BSTS make up more than 20% of cancer arising in WS patients.[17] Soft tissue sarcomas that have been identified in WS patients include malignant fibrous histiocytoma, malignant peripheral nerve sheath tumor, fibrosarcoma, rhabdomyosarcoma, liposarcoma, and synovial sarcoma. The reason for the high representation of mesenchymal tumors in WS patients has not been clarified.

WRN encodes a protein with 1432 amino acids that possesses both 3′→5′ DNA helicase and 3′→5′ DNA exonuclease activities. DNA helicases are enzymes that unwind the energetically stable double-stranded structure of DNA to provide a single-stranded template for important cellular processes such as replication, base excision repair, homologous recombination, and telomere maintenance.[15,18,19]

Several epidemiological studies have already been carried out on WRN-Cys1367Arg. Most of those studies focused on the diseases relating to the WS phenotype, such as myocardial infarction, diabetes mellitus and lymphomas. The more frequent Cys1367 allele has been reported to be associated with a lower frequency of osteoporosis in postmenopausal Japanese women,[20] whereas the minor allele Arg1367 may be associated with a lower risk of myocardial infarction and type 2 diabetes mellitus in the Japanese population.[21,22] With regard to cancer risk, Arg1367 was reported to be associated with a decreased risk of non-Hodgkin lymphoma among women in Connecticut.[23] Of note, in three out of the four reports listed above, WRN-Cys1367Arg showed protective effects against the various diseases associated with WS phenotype, and the same held true for our observations on BSTS.

Based on the observations that all of the pathogenic *WRN* mutations identified so far result in truncation of the C-terminal

of the WRN protein, it has been proposed that a lack of the C-terminal nuclear localization signal is important in the pathogenesis of WS.[24,25] Although WRN-Cys1367Arg is located adjacent to the nuclear localization signal, a previous report has failed to detect any significant difference between WRN (Arg1367) and WRN (Cys1367) with respect to their nuclear localization,[26] or helicase and helicase-coupled exonuclease activity.[27] Other possible explanations for the observed association include the allelic difference of WRN-Cys1367Arg in the interactions with other proteins or the presence of unknown functionally responsible polymorphisms that are in linkage disequilibrium with WRN-Cys1367Arg.

Some of the SNPs on the genes responsible for a hereditary form of cancer have shown association with a sporadic form of the same type of cancer.[28–31] Our findings on the *WRN* SNP on BSTS may add another example of the possible sharing, at least in part, of the oncogenesis pathway between the monogenic and polygenic forms of the same type of cancer and a continuity of the genotype-phenotype spectrum.

It may be controversial to analyze the possible genetic backgrounds with all types of BSTS combined because of its highly heterogeneous nature in histology. However, little if any information is currently available on the genetic predisposition to the sporadic forms of BSTS, and genetic factors that are common to most BSTSs may exist, as well as those specific to certain subgroups. As one of the first exploratory analyses on the genetic susceptibility of sporadic BSTS, the primary role of this study is to generate hypotheses, which deserve further validation. To validate the hypothesis, evidence should be sought both in statistical replication and meta-analysis using future case-control panels and also in biological functional analyses.

Acknowledgments

We are grateful to Dr Matsuhiko Hayashi and Dr Yoichi Ohno of Department of Internal Medicine, Keio University School of Medicine, and Dr Fumihiko Tanioka of Department of Pathology and Laboratory Medicine, Iwata Municipal General Hospital for their help in collecting blood samples, and Ms. Sachiyo Mimaki and Mr Hirohiko Totsuka of the Center for Medical Genomics, National Cancer Center for providing considerable contributions to DNA sequencing and statistical analysis. This study was supported by the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio) and in part by Grant-in-Aid for Scientific Research on Priority Area (18014009) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1 Dorfman HD, Czerniak B. Bone cancers. *Cancer* 1995; **75**: 203–10.
2 Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: an analysis of 26 758 cases. *Int J Cancer* 2006; **119**: 2922–30.
3 Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin* 2007; **57**: 43–66.
4 Ross JA, Severson RK, Davis S, Brooks JJ. Trends in the incidence of soft tissue sarcomas in the United States from 1973 through 1987. *Cancer* 1993; **72**: 486–90.
5 Zahm SH, Fraumeni JF Jr. The epidemiology of soft tissue sarcoma. *Semin Oncol* 1997; **24**: 504–14.
6 Polednak AP. Primary bone cancer incidence in black and white residents of New York State. *Cancer* 1985; **55**: 2883–8.
7 Mark RJ, Poen J, Tran LM, Fu YS, Selch MT, Parker RG. Postirradiation sarcomas. A single-institution study and review of the literature. *Cancer* 1994; **73**: 2653–62.
8 Sakiyama T, Kohno T, Mimaki S *et al.* Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *Int J Cancer* 2005; **114**: 730–7.
9 Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 1513–30.
10 Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38**: 209–13.
11 Sandberg AA. Cytogenetics and molecular genetics of bone and soft-tissue tumors. *Am J Med Genet* 2002; **115**: 189–93.
12 Alderborn A, Kristofferson A, Hammerling U. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res* 2000; **10**: 1249–58.
13 Breslow NE, Day NE. Statistical methods in cancer research. Vol. I. the analysis of case-control studies. *IARC Sci Pub* 1980: 5–338.
14 Holm S. A simple sequentially rejective multiple test procedure. *Scand J Statistics* 1979; **6**: 65–70.
15 Yu CE, Oshima J, Fu YH *et al.* Positional cloning of the Werner's syndrome gene. *Science* 1996; **272**: 258–62.
16 Epstein CJ, Martin GM, Schultz AL, Motulsky AG. Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. *Medicine (Baltimore)* 1966; **45**: 177–221.
17 Goto M, Miller RW, Ishikawa Y, Sugano H. Excess of rare cancers in Werner syndrome (adult progeria). *Cancer Epidemiol Biomarkers Prev* 1996; **5**: 239–46.
18 Gray MD, Shen JC, Kamath-Loeb AS *et al.* The Werner syndrome protein is a DNA helicase. *Nat Genet* 1997; **17**: 100–3.
19 Kamath-Loeb AS, Shen JC, Loeb LA, Fry M. Werner syndrome protein. II. Characterization of the integral 3′→5′ DNA exonuclease. *J Biol Chem* 1998; **273**: 34 145–50.
20 Ogata N, Shiraki M, Hosoi T, Koshizuka Y, Nakamura K, Kawaguchi H. A polymorphic variant at the Werner helicase (WRN) gene is associated with bone density, but not spondylosis, in postmenopausal women. *J Bone Miner Metab* 2001; **19**: 296–301.
21 Ye L, Miki T, Nakura J, Oshima J *et al.* Association of a polymorphic variant of the Werner helicase gene with myocardial infarction in a Japanese population. *Am J Med Genet* 1997; **68**: 494–8.
22 Hirai M, Suzuki S, Hinokio Y *et al.* WRN gene 1367 Arg allele protects against development of type 2 diabetes mellitus. *Diabetes Res Clin Pract* 2005; **69**: 287–92.
23 Shen M, Zheng T, Lan Q *et al.* Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma among women in Connecticut. *Hum Genet* 2006; **119**: 659–68.
24 Matsumoto T, Shimamoto A, Goto M, Furuichi Y. Impaired nuclear localization of defective DNA helicases in Werner's syndrome. *Nat Genet* 1997; **16**: 335–6.
25 Yu CE, Oshima J, Wijsman EM *et al.* Mutations in the consensus helicase domains of the Werner syndrome gene. Werner's Syndrome Collaborative Group. *Am J Hum Genet* 1997; **60**: 330–41.
26 Bohr VA, Metter EJ, Harrigan JA *et al.* Werner syndrome protein 1367 variants and disposition towards coronary artery disease in Caucasian patients. *Mech Ageing Dev* 2004; **125**: 491–6.
27 Kamath-Loeb AS, Welcsh P, Waite M, Adman ET, Loeb LA. The enzymatic activities of the Werner syndrome protein are disabled by the amino acid polymorphism R834C. *J Biol Chem* 2004; **279**: 55 499–505.
28 Healey CS, Dunning AM, Teare MD *et al.* A common variant in BRCA2 is associated with both breast cancer risk and prenatal viability. *Nat Genet* 2000; **26**: 362–4.
29 Slattery ML, Samowitz W, Ballard L, Schaffer D, Leppert M, Potter JD. A molecular variant of the APC gene at codon 1822: its association with diet, lifestyle, and risk of colon cancer. *Cancer Res* 2001; **61**: 1000–4.
30 Freedman ML, Penney KL, Stram DO, Le Marchand L *et al.* Common variation in BRCA2 and breast cancer risk: a haplotype-based analysis in the Multiethnic Cohort. *Hum Mol Genet* 2004; **13**: 2431–41.
31 Freedman ML, Penney KL, Stram DO *et al.* A haplotype-based case-control study of BRCA1 and sporadic breast cancer risk. *Cancer Res* 2005; **65**: 7516–22.

doi: 10.1111/j.1349-7006.2007.00692.x

Supplementary Material

The following supplementary material is available for this article:

Fig. S1. Age distributions of cases and controls in this study. The case appears to have two peaks, before and after around 40 years of age.

A Monte Carlo simulation experiment to compare family wise error rate for the multiple testing correction by six single nucleotide polymorphisms (SNPs) (six hypotheses examined for the joint analysis) and that by 50 SNPs (initial candidate SNPs screened in the first screening). As the framework of simulation, we set the following conditions:

Condition 1.
The total number of SNPs to be examined is set as $m = 50$ and the number of 'true' disease-associated SNPs (positive SNPs) among the 50 SNPs as $N_p = 3$ or 5.

Condition 2.
The population allelic odds ratio for the disease-associated N_p SNPs is $\psi = 1.3$, 1.5 or 1.7, while the population odds ratio for the remaining m-N_p SNPs unrelated to the disease is 1.0.

Condition 3.
In the first stage, sample size is set as 240 cases and 685 control subjects. The second stage sample size is 304 cases and 834 control subjects. In a joint analysis, sample size is set as 544 cases and 989 control subjects.

Condition 4.
The proportion of allele X in the control population is a random variable uniformly distributed in unit interval (0.05, 0.95).

Condition 5.
The criteria to evaluate the performance of each method are two indicators, sensitivity and specificity in the joint analysis.

Condition 6.
The Monte-Carlo simulation to calculate sensitivity and specificity is repeated 10 000 times, and the mean values of indicators are calculated.

Table S1 shows a summary result of the simulation experiment, which suggests that our study was designed to contain the overall false-positive (type 1 error) rate (1-specificity) to 5% or 8% at the power of 90%, if there are five or three true SNPs, respectively, included in our starting 50 candidate SNPs. On the other hand, the false-positive rates for the multiple testing correction by the 50 SNPs with OR = 1.5 or 1.7 are less than 1%.

Table S1. Simulation results for each multiple testing correction method

Table S2. Statistics of subgroup analysis by recessive and dominant models on all the six single nucleotide polymorphisms (SNPs) analyzed in the second stage of genotyping

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1349-7006.2008.00692.x
<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1349-7006.2008.00692.x>
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Pfetin as a Prognostic Biomarker of Gastrointestinal Stromal Tumors Revealed by Proteomics

Yoshiyuki Suehara,[1,7,8] Tadashi Kondo,[1] Kunihiko Seki,[4] Tatsuhiro Shibata,[2] Kiyonaga Fujii,[1] Masahiro Gotoh,[3] Tadashi Hasegawa,[4] Yasuhiro Shimada,[5] Mitsuru Sasako,[6] Tadakazu Shimoda,[4] Hisashi Kurosawa,[8] Yasuo Beppu,[7] Akira Kawai,[7] and Setsuo Hirohashi[1]

Abstract **Purpose:** We aimed to develop prognostic biomarkers for gastrointestinal stromal tumors (GIST) using a proteomic approach.

Experimental Design: We examined the proteomic profile of GISTs using two-dimensional difference gel electrophoresis. The prognostic performance of biomarker candidates was examined using a large-scale sample set and specific antibodies.

Results: We identified 43 protein spots whose intensity was statistically different between GISTs with good and poor prognosis. Mass spectrometric protein identification showed that the 43 spots corresponded to 25 distinct gene products. Eight of the 43 spots derived from pfetin, a potassium channel protein, and four of the eight pfetin spots had a high discriminative power between the two groups. Western blotting and real-time PCR showed that pfetin expression and tumor metastasis were inversely related. The prognostic performance of pfetin was also examined by immunohistochemistry on 210 GIST cases. The 5-year metastasis-free survival rate was 93.9% and 36.2% for patients with pfetin-positive and pfetin-negative tumors, respectively ($P < 0.0001$). Univariate and multivariate analyses revealed that pfetin expression was a powerful prognostic factor among the clinicopathologic variables examined, including risk classification and c-kit– or platelet-derived growth factor receptor A mutation status.

Conclusions: These results establish pfetin as a powerful prognostic marker for GISTs and may provide novel therapeutic strategies to prevent metastasis of GIST.

Gastrointestinal stromal tumors (GIST) are the most common primary mesenchymal tumors of the digestive tract, with a prevalence of 15 to 20 per 1,000,000 (1, 2). The clinical course of GISTs spans a wide spectrum from a curable disorder to a highly malignant disease that leads to metastasis and death. Thus, the molecular background of GISTs has been extensively studied to predict the behavior of individual tumors. GISTs are characterized by the presence of mutations and overexpression of c-kit and, clinically, by their response to treatment with imatinib (3–7). Tumor location and certain molecular aberrations, including c-kit, platelet-derived growth factor (PDGFR), and p16 alterations, have been found to correlate with patient prognosis and response to treatment with imatinib (8–12). However, none of these variables has been proven to be clinically useful in improving patient outcomes yet. GISTs arise from the intestinal cells of Cajal, which are the mesenchymal pacemaker cells of the guts (13), their biological characteristics, however, remain largely obscure.

Recent comprehensive studies offered a global view of molecular aberrations associated with the malignant spectrum of GISTs. Genomic studies using fluorescence *in situ* hybridization and array-based comparative genomic hybridization identified chromosomal regions frequently amplified and target genes within these regions, the copy number status of which correlated with tumor behavior (14, 15). Global mRNA expression studies using DNA microarrays identified the genes that are involved in the signaling pathways specific to kit or PDGFR and aberrantly regulated in GISTs (16), the genes associated with histologic features denoting malignancy (17), and the genes differentially expressed based on the KIT genotype and GIST anatomic site (18). These comprehensive studies will further increase our understanding of the biology of GIST and lead to the

Authors' Affiliations: [1]Proteome Bioinformatics Project, [2]Cancer Genomics Project, and [3]Pathology Division, National Cancer Center Research Institute; [4]Clinical Laboratory Division, [5]Gastrointestinal Oncology Division, [6]Gastric Surgery Division, and [7]Orthopedic Surgery Division, National Cancer Center Hospital; and [8]Department of Orthopedic Surgery, Juntendo University School of Medicine, Tokyo, Japan

Received 6/15/07; revised 12/25/07; accepted 1/9/08.

Grant support: Ministry of Health, Labor, and Welfare and by the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation of Japan (patent no. 2006-286087).

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

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Current address for T. Hasegawa: Department of Clinical Pathology, Sapporo Medical University School of Medicine.

Requests for reprints: Tadashi Kondo, Proteome Bioinformatics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. Phone: 81-3-3542-2511, ext. 3004; Fax: 81-3-357-5298; E-mail: takondo@gan2.res.ncc.go.jp.

doi:10.1158/1078-0432.CCR-07-1478

Table 1. Clinicopathologic features of the training set samples (17 cases)

Sample No	Age	Gender	Site	Histologic types	Size (cm)	Risk classification*
1	68	M	Stomach	Spindle	19	High
2	56	M	Stomach	Spindle	38	High
3	58	M	Stomach	Spindle	13	High
4	50	F	Rectum	Spindle	4	High
5	51	M	Stomach	Mixed (spindle main)	12	High
6	34	M	Small intestine	Spindle	18	High
7	68	M	Small intestine	Mixed (spindle main)	7	High
8	72	F	Stomach	Spindle	25	High
9	64	F	Stomach	Spindle	3.5	Low
10	64	F	Stomach	Spindle	4	Low
11	54	M	Stomach	Spindle	10	Intermediate
12	68	M	Small intestine	Spindle	3.7	Low
13	77	M	Small intestine	Spindle	4	Low
14	40	F	Stomach	Spindle	10	Intermediate
15	52	M	Stomach	Spindle	7	Intermediate
16	76	M	Small intestine	Spindle	7	Intermediate
17	81	M	Stomach	Spindle	5.5	Intermediate

NOTE: PDGFR mutations: All samples lacked of PDGFR mutations. Detail data: Supplementary Table S1.
Abbreviations: NED, no evidence of disease; AWD: alive with disease; DOD, dead of disease.
*Prognostic classification based on tumor size and MIB-1 grade (Hasegawa, T. et al. Hum Pathol. 2002, 33:669-676).

development of practical tumor markers to support individualized therapy (8). Emerging technologies that examine the overall features of the expressed proteins, namely the proteome, have identified many candidate proteins associated with early diagnosis (19), differential diagnosis (20), prognosis (21), and response to chemotherapy (22) in various diseases. Many lines of evidence have indicated that DNA copy number and mRNA expression levels do not necessarily correspond to the protein contents, and that posttranslational modifications cannot be predicted by DNA sequences (23, 24), suggesting that proteomic studies offer unique data that cannot be obtained by other approaches. The proteomic profile of GISTs has not been established yet, and a proteomic study using a large-scale clinical sample set would complement the genome and transcriptome studies.

In this report, we did a comprehensive quantitative expression study on the intact proteins of GIST clinical samples using two-dimensional difference gel electrophoresis and mass spectrometry. Proteomic studies on peptides have been used to develop tumor markers, but intact proteins have not been considered for this purpose, with a few exceptions. Two-dimensional difference gel electrophoresis, as the most advanced form of two-dimensional gel electrophoresis, has the great advantage of being able to be used to study intact proteins. We found that the expression levels of 43 proteins, including eight variants of pfetin (predominantly fetal-expressed tetramerization domain; potassium channel tetramerisation domain containing protein 12), which was originally reported as a protein highly expressed in fetal cochlea and brain (25), correlated with prognosis. We verified the prognostic value of pfetin expression on 210 GIST cases using immunohistochemistry. Our findings indicate that the use of pfetin expression as a prognostic indicator may facilitate tailored medical care for GIST patients.

Materials and Methods

Patients and clinical information. We examined the tumor tissues of 212 GIST patients who underwent surgery at the National Cancer Center Hospital consecutively from October 1977 to December 2005. All patients underwent resection with curative intent and were not treated with adjuvant chemotherapy, including treatment with imatinib, until distant metastasis was diagnosed. Histologic features of the tissues were reviewed by three board-certified pathologists (K.S., T.S., and T.H.). Diagnosis was based on the WHO classification system for soft-tissue tumors (26), including the examination of tumor size, presence of necrosis, differentiation, mitotic rate, MIB-1 index, presence of epithelioid cells, and CD34 and CD117 expression. Using this large, well-characterized single hospital–based sample set, we were able to identify proteomic features that differ significantly when examined in relation to certain clinicopathologic variables. This project was approved by the institutional review board of the National Cancer Center.

Previous reports indicated that GIST patients that were histologically classified as of being at low or intermediate risk did not develop metastases within 2 y postsurgery, whereas GIST patients histologically classified as of being at high risk developed metastases within 1 y postsurgery (27). For proteomic analysis, we grouped the GIST samples into two groups. GISTs that had metastases at diagnosis or developed metastases within 1 y postsurgery and were categorized in the high-risk group based on their histologic features were defined as poor-prognosis GISTs (P-GIST; Table 1, samples 1-8). GISTs that did not have metastases within 2 y postsurgery and were grouped in the low- or intermediate-risk group based on the histologic features were defined as good-prognosis GISTs (G-GIST; Table 1, samples 9-17). The samples listed in Supplementary Table S1 were excluded from this classification; samples 18, 19, 24, and 25 were excluded because RNA data were not available for the validation study and the other samples because they did not meet the criteria for classification either as P-GISTs or G-GISTs.

For the immunohistochemical study, we selected 210 patients who did not have distant metastases at the time of surgery.

Protein expression profiling. Frozen samples were crushed to powder with a CryoPress (Microtech Nichion) under cooling with liquid nitrogen. The frozen powder was then treated with urea lysis buffer

Table 1. Clinicopathologic features of the training set samples (17 cases) (Cont'd)

Type of KIT mutation	Metastatic site (first development)	Metastasis time after diagnosis (mo)	Follow-up time after diagnosis (mo)	Follow-up status
Wild-type	Peritoneal metastasis	8	16	DOD
EX11 deletion	Peritoneal metastasis	7	9	DOD
EX11 deletion	Peritoneal metastasis	6	11	DOD
EX11 deletion	Peritoneal metastasis	11	60	AWD
EX11 deletion	Peritoneal metastasis	5	69	AWD
EX11 deletion	Liver metastasis	At diagnosis	31	AWD
EX9 insertion	Peritoneal metastasis	At diagnosis	9	AWD
EX11 559 V-D	Peritoneal metastasis	5	8	AWD
Wild-type	—	—	68	NED
Wild-type	—	—	81	NED
EX11 560 V-G	—	—	77	NED
EX9 insertion	—	—	50	NED
Wild-type	—	—	69	NED
EX11 559 V-D	—	—	88	NED
EX11 576 L-P	—	—	62	NED
EX11 deletion	—	—	48	NED
EX11 559 V-D	—	—	43	NED

(6 mol/L urea, 2 mol/L thiourea, 3% CHAPS, and 1% Triton X-100). After centrifugation at 15,000 rpm for 30 min, the supernatant was used as the source of cellular proteins for protein expression studies.

Two-dimensional difference gel electrophoresis was done as described previously (20, 28, 29). In brief, the internal control sample was prepared by mixing a portion of all individual samples. Five micrograms of the internal control sample and of each individual sample were labeled with Cy3 and Cy5, respectively (CyDye DIGE Fluor saturation dye; GE Healthcare Biosciences) according to the manufacturer's instructions. The differently labeled protein samples were mixed and separated by two-dimensional gel electrophoresis, which was achieved using IPG DryStrip gels for the first dimension separation (length, 24 cm; isoelectric point range, between 4 and 7; GE Healthcare Biosciences) and SDS-PAGE for the second dimension separation (EttanDalt II; GE Healthcare Biosciences). The gels were scanned using laser scanners (Tyhoon Trio; GE Healthcare Biosciences) at appropriate wavelengths. For all spots, the intensity of the Cy5 image was normalized by that of the Cy3 image in the identical gel so that gel-to-gel differences were compensated, using the DeCyder image software (GE Healthcare Biosciences). System reproducibility was verified by comparing the protein profiles obtained from three independent separations of the identical sample (sample 22; Supplementary Table S1). Scatter plot analysis revealed that the standardized intensity of >96% of the spots ranged within a 2.0-fold difference (Supplementary Fig. S1). Representative two-dimensional images with the numbers of the identified spots are shown in Fig. 1A and Supplementary Fig. S3.

Data analysis. We identified protein spots whose intensity was statistically (Wilcoxon test, $P < 0.001$) different between the groups examined. Hierarchical clustering, principal component analysis, correlation matrix study, and spot ranking were done using the Expressionist software (Genedata).

Protein identification by mass spectrometry. Proteins corresponding to the spots of interest were identified by mass spectrometry according to our previous report (20, 30). Cy5-labeled proteins separated by 2D-PAGE were recovered in gel plugs and digested with modified trypsin (Promega). The trypsin digests were subjected to liquid chromatography coupled with tandem mass spectrometry equipped with a nanoelectrospray ion source (Paradigm MS4 dual solvent delivery system; Michrom BioResources, Inc.) for microflow high performance liquid chromatography, an HTS PAL auto sampler (CTC Analytics), and a Finnigan LTQ linear ion trap mass spectrometer (Thermo Electron Co.) equipped with a nanoelectrospray ion source (AMR, Inc.). The Mascot software (version 2.1; Matrix Science) was used to search for the mass of the peptide ion peaks against the SWISS-PROT database (Homo sapiens, 12867 sequence in Sprot_47.8 fasta file). Proteins with a Mascot score of 35 or more were used for protein identification. When multiple proteins were identified in a single spot, the proteins with the highest number of peptides were considered as those corresponding to the spot.

Western blotting and immunohistochemistry. Protein samples were separated by SDS-PAGE and subsequently blotted on a nitrocellulose membrane. The membrane was incubated with rabbit polyclonal antibody against pfetin (1:1,000 dilution) kindly provided by Dr. Morton (25), and then horseradish peroxidase-conjugated secondary antibody (1:1,000 dilution; GE Healthcare Biosciences). Pfetin was detected using an emhanced chemiluminescence system (GE Healthcare Biosciences) and LA 1000 (Fuji film).

Pfetin expression was examined immunohistochemically using paraffin-embedded tissues. In brief, 4-μm-thick tissue sections were autoclaved in 10 mmol/L citrate buffer (pH 6.0) at 121°C for 30 min and incubated with the antibody against pfetin (1:500 dilution). Immunostaining was done according to the streptavidin-biotin peroxidase method using the Strept ABC Complex/horseradish peroxidase kit (DAKO). One pathologist (K. S.) and one medical doctor (Y. S.) reviewed the sections stained with antipfetin antibody in a blinded fashion regarding clinical data (age, sex, anatomic site, and outcome). Positively stained cells were defined as those that had higher staining intensity than that of vascular endothelial cells, which served as positive controls. Cases with >20% of tumor cells stained positively with the antipfetin antibody were considered as pfetin positive, whereas cases with <20% pfetin-positive tumor cells were considered as pfetin negative. In most cases, the difference was so obvious that two reviewers had consistent results. Inconsistencies, if any, were resolved by discussion, as a usual process of pathologic diagnosis in the hospital.

Mutation study for c-kit and PDGFRA. We examined the c-kit and PDGFRA genes for the presence of mutations as previously described (31) in the 39 cases where DNA samples were available. In brief, DNA was extracted from the frozen tissues, and the exons including the frequent mutation sites for c-kit and PDGFRA were amplified by PCR. The PCR products were purified with 2% agarose gel electrophoresis, extracted with an QIAquick PCR Purification kit (Qiagen), and sequenced using an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). The primer sets for c-kit were as follows: 5′-TCTAGTGCATTCAAGCACAATGG-3′ and 5′-CATGACTGATATGGTAGACAGAG-3′ for exon 9, and 5′-CCAGAGTGCTCTAATGACTGAGAC-3′ and 5′-AAAGGTGACATGGAAAGCCCCTG-3′ for exon 11.The primer set for exon 12 of the PDGFRA gene was 5′-CCTGGTCATTTATAGAAACCGAG-3′

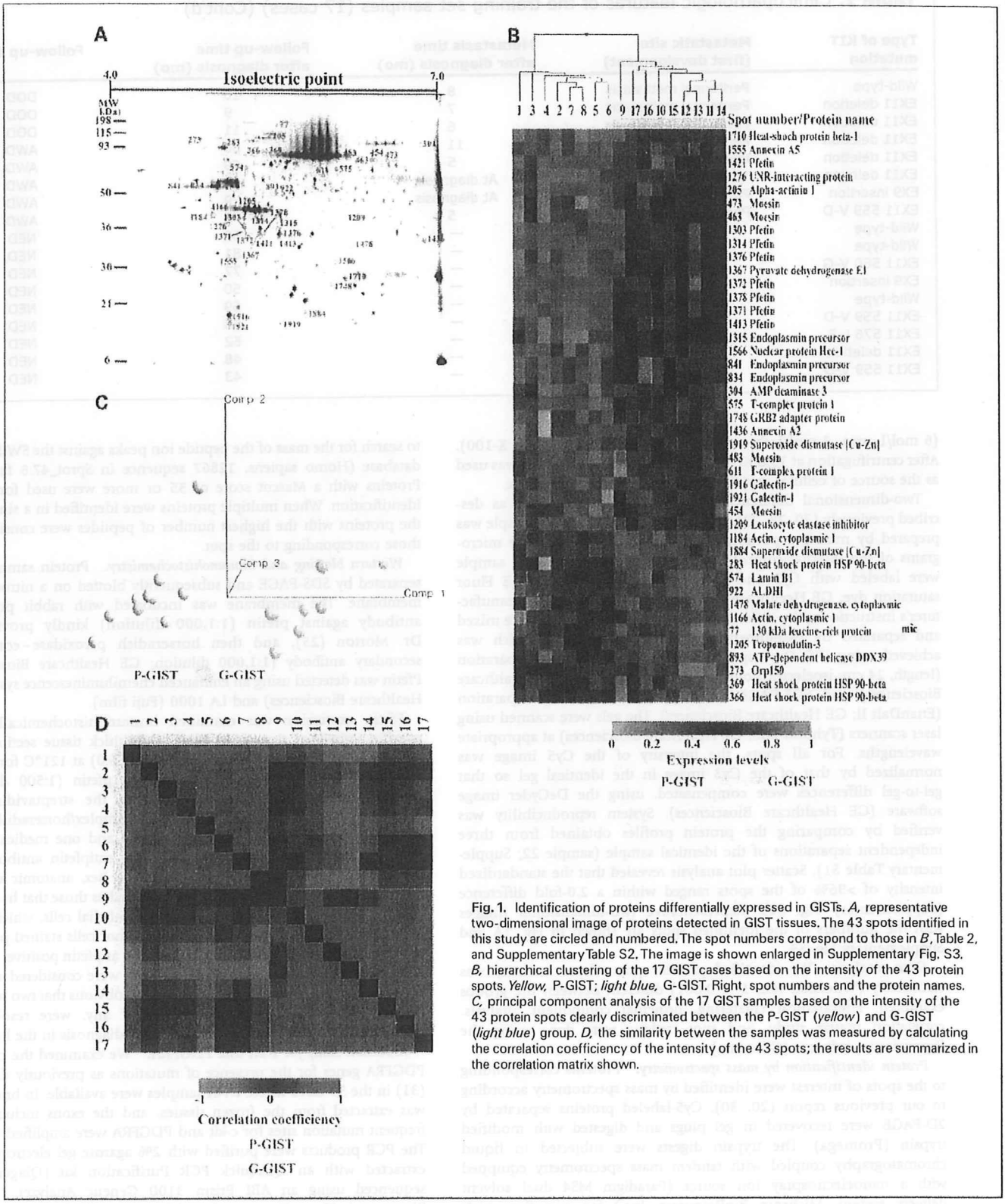

Fig. 1. Identification of proteins differentially expressed in GISTs. *A,* representative two-dimensional image of proteins detected in GIST tissues. The 43 spots identified in this study are circled and numbered. The spot numbers correspond to those in *B*, Table 2, and Supplementary Table S2. The image is shown enlarged in Supplementary Fig. S3. *B,* hierarchical clustering of the 17 GIST cases based on the intensity of the 43 protein spots. *Yellow,* P-GIST; *light blue,* G-GIST. Right, spot numbers and the protein names. *C,* principal component analysis of the 17 GIST samples based on the intensity of the 43 protein spots clearly discriminated between the P-GIST (*yellow*) and G-GIST (*light blue*) group. *D,* the similarity between the samples was measured by calculating the correlation coefficiency of the intensity of the 43 spots; the results are summarized in the correlation matrix shown.

and 5′-CTCCCATCTTGAGTCATAAGGCA-3′. PCR cycling variables were as follows: cycle at 96°C for 1 min; 50 cycles at 94°C for 30 s, 56°C for 30 s, and 68°C for 2.5 min; and finally 1 cycle at 72°C for 5 min.

Quantitative reverse transcription-PCR. We extracted mRNA and generated cDNA using the Super Script III kit (Invitrogen) in the 39 cases where mRNA samples were available. The quantitative amplification was monitored with Taq Man Gene Expression Assays

Table 2. A list of identified protein

Spots no*	Accession no†	Identified protein†	Wilcoxon test P	Fold difference (ratio of means)	Overall rank
1378	Q96CX2	Potassium channel tetramerisation domain containing protein 12 (Pfetin)	8.23E-05	4.862	1
1314	Q96CX2	Potassium channel tetramerisation domain containing protein 12 (Pfetin)	8.23E-05	5.805	2
1,371	Q96CX2	Potassium channel tetramerisation domain containing protein 12 (Pfetin)	8.23E-05	3.781	5
1413	Q96CX2	Potassium channel tetramerisation domain containing protein 12 (Pfetin)	1.65E-04	3.198	26
575	P17987	T-complex protein 1, α subunit	3.29E-04	1.405	28
893	O00148	ATP-dependent helicase DDX39	6.99E-04	2.321	41
1367	P11177	Pyruvate dehydrogenase estrone component β subunit	8.94E-04	3.669	24
1303	Q96CX2	Pfetin	9.87E-04	2.917	4
1919	P00441	Superoxide dismutase (Cu-Zn)	1.33E-03	2.547	12
611	P17987	T-complex protein 1, α subunit	1.56E-03	1.982	9
1184	P60709	Actin, cytoplasmic 1	1.56E-03	1.902	20
1376	Q96CX2	Pfetin	1.56E-03	2.850	38
834	P14625	Endoplasmin precursor	1.75E-03	3.487	8
483	P26038	Moesin	1.86E-03	1.558	15
369	P08238	Heat shock protein HSP 90-β	1.86E-03	1.977	16
922	P05091	Aldehyde dehydrogenase, mitochondrial precursor	2.04E-03	1.359	31
273	Q9Y4L1	150 kDa oxygen-regulated protein precursor	3.32E-03	1.384	36
1436	P07355	Annexin A2	3.32E-03	1.590	39
1,748	P62993	Growth factor receptor-bound protein 2	3.32E-03	1.545	40
205	P12814	α-actinin 1	3.70E-03	1.993	11
1710	P04792	Heat-shock protein β-1	3.70E-03	2.078	13
574	P20700	Lamin B1	3.70E-03	2.575	14
1555	P08758	Annexin A5	3.70E-03	1.438	19
454	P26038	Moesin	3.70E-03	1.441	37
1315	P14625	Endoplasmin precursor	3.73E-03	2.895	23
1276	Q9Y3F4	Serine-threonine kinase receptor-associated protein	3.85E-03	1.650	29
304	Q01432	AMP deaminase 3	4.00E-03	1.691	22
1372	Q96CX2	Pfetin	4.04E-03	2.972	30
1478	P40925	Malate dehydrogenase, cytoplasmic	4.08E-03	1.534	27
1916	P09382	Galectin-1	4.21E-03	1.917	7
1209	P30740	Leukocyte elastase inhibitor	4.33E-03	2.380	18
1884	P00441	Superoxide dismutase (Cu-Zn)	5.25E-03	3.121	3
77	P42704	130 kDa leucine-rich protein	5.51E-03	1.456	21
1166	P60709	Actin, cytoplasmic 1	5.51E-03	3.971	42
1921	P09382	Galectin-1	6.22E-03	1.912	6
1421	Q96CX2	Pfetin	6.22E-03	1.665	34
366	P08238	Heat shock protein HSP 90-β	7.02E-03	1.705	25
1566	P82979	Nuclear protein Hcc-1	7.02E-03	1.526	43
473	P26038	Moesin	7.59E-03	1.323	32
1205	Q9NYL9	Tropomodulin-3	7.90E-03	1.935	10
283	P08238	Heat shock protein HSP 90-β	7.90E-03	1.300	33
463	P26038	Moesin	7.90E-03	1.524	35
841	P14625	Endoplasmin precursor	9.45E-03	2.478	17

*Spot numbers refer to those in Fig. 1A and Supplementary Fig. S3.
†Accession numbers of protein were derived from SWISS-PROT and National Center for Biotechnology Information nonredundant databases.
‡Observed isoelectric point and molecular weight calculated according to location on the two-dimensional gel.
§Theoretical isoelectric point and molecular weight obtained from Swiss-Prot and the ExPASy database. (http://au.expasy.org).
‖Mascot score for the identified proteins based on the peptide ions score ($P < 0.05$; http://www.matrixscience.com).

using premade primers for pfetin and Human glyceraldehyde-3-phosphate dehydrogenase, and Taq Man Universal PCR Master Mix with 7500 Real-time PCR system (Applied Biosystems) according to the manufacturer's instructions.

Statistical analysis. The tumor-specific and metastasis-free survival times were calculated from the first resection of the primary tumor to death from tumor-specific causes or to first evidence of metastasis, respectively. All time-to-event end points were computed by the Kaplan-Meier method (32). Patients who died of causes unrelated to GISTs were censored at the time of death. Potential prognostic factors were identified by univariate analysis using the log-rank test. Independent prognostic factors were evaluated using a Cox's proportional hazards regression model and a stepwise selection procedure. To arrive at a parsimonious multivariate model, covariates were selected into the

Table 2. A list of identified protein (Cont'd)

pI (obs)‡	pI (cal)§	MW (obs; kDa)‡	MW (cal; kD)§	Protein score‖	Peptide matches	Sequence coverage (%)
6.4	5.5	35.1	35.7	406	8	20.0
5.4	5.5	44.0	35.7	338	7	19.7
6.2	5.5	35.7	35.7	146	2	8.3
6.5	5.5	34.5	35.7	306	8	23.4
5.1	5.8	73.3	60.3	267	6	12.4
5.5	5.5	51.8	49.1	109	2	3.7
6.4	6.2	35.1	39.2	174	3	9.7
5.2	5.5	44.7	35.7	236	5	14.5
6.6	5.7	15.3	15.8	60	1	7.8
4.9	5.8	67.9	60.3	667	11	22.5
5.9	5.3	46.3	41.7	276	4	11.2
6.5	5.5	35.1	35.7	396	11	23.7
4.7	4.8	57.2	92.5	977	18	20.5
5.2	6.1	84.0	67.7	208	4	6.9
4.4	5.0	89.4	83.1	263	5	7.3
5.6	6.6	50.0	56.4	323	5	10.1
4.7	5.2	113.7	111.3	102	2	2.4
6.9	7.6	34.5	38.5	262	5	13.3
6.0	5.9	25.5	25.2	79	2	7.4
4.4	5.3	114.1	103.1	593	12	12.9
6.1	6.0	27.5	22.8	491	15	54.1
4.2	5.1	75.1	66.3	1616	33	36.4
6.1	4.9	31.5	35.8	96	2	5.3
5.4	6.1	82.3	67.7	366	8	9.2
5.5	4.8	44.3	92.5	216	5	4.7
5.1	5.0	45.0	38.4	250	5	15.4
4.8	6.5	113.4	88.8	185	3	4.6
6.3	5.5	35.1	35.7	118	2	8.3
6.2	6.9	33.6	36.3	116	2	5.7
6.1	5.3	16.9	14.6	244	6	35.8
5.2	5.9	46.0	42.7	551	9	23.0
6.8	5.7	18.4	15.8	139	2	17.0
4.5	5.5	195.2	145.2	120	2	1.7
5.9	5.3	46.3	41.7	246	7	15.5
6.1	5.3	15.8	14.6	90	3	11.9
6.4	5.5	34.5	35.7	336	13	23.4
4.4	5.0	89.4	83.1	420	8	13.8
6.1	6.1	30.6	23.5	196	2	13.9
5.6	6.1	80.5	67.7	775	15	18.8
5.2	5.1	46.0	39.6	168	3	9.1
4.1	5.0	113.1	83.1	907	17	22.1
5.4	6.1	82.3	67.7	356	7	9.7
4.7	4.8	57.2	92.5	159	3	4.0

model only if they contributed significantly to the fit of the model based on the χ^2 test value. *P* value differences of <0.05 were considered to be significant. Statistical analyses were done using the SPSS statistical package (SPSS).

Results

We compared the protein expression profiles between nine G-GISTs and eight P-GISTs using two-dimensional difference gel electrophoresis. We selected 1,513 protein spots that appeared in at least 75% of the images of the Cy3-labeled internal control sample to decrease irrelevant expression data. The G- and P-GIST samples were not classified into their respective groups based on the overall protein expression features of the samples (Supplementary Fig. S2). However, we found that 43 protein spots had significantly different intensity between the two groups ($P < 0.01$). The localization of the 43 spots on the two-dimensional image is shown in Fig. 1A (enlarged image in Supplementary Fig. S3). Hierarchical clustering and principal component analysis accurately classified the 17 GIST samples into either the G- or the P-GIST group based on the intensity of the 43 selected spots (Fig. 1B and C). The profiles of the 43 spots were similar between samples of the same group and different between

the two groups (Fig. 1D). Mass spectrometric protein identification revealed that the 43 protein spots corresponded to 25 distinct gene products (Fig. 1B; Table 2; Supplementary Table S2).

We aimed to prioritize the protein spots according to their discriminative power for the two groups. We created a classifier based on a support vector machine algorithm that used the intensity of the 43 spots and ranked the 43 spots according to their contribution to the classification using support vector machine weight algorithm (Table 2). We found that 4 of the 8 identified pfetin spots were ranked within the top 10 protein spots whose intensity was different between the groups (Table 2), and that pfetin spots appeared 8 times in the list of the 43 protein spots (Fig. 1B; Table 2).

Pfetin is highly expressed in fetal cochlea and brain (25), consistent with the fact that the origin of GIST is Cajal cells, neuronal cells in the gut. Thus, we further validated the relationship of the expression of pfetin with the malignant potential of GISTs. SDS-PAGE/Western blotting showed that the expression of pfetin was lower in the P-GIST compared with the G-GIST group (Fig. 2A). Two bands on SDS-PAGE/Western blotting corresponded to the location of protein spots for pfetin variants on the two-dimensional image (Fig. 1A; Supplementary Fig. S3). These results were further validated in an additional four GIST samples that were not included in the initial proteome study (Supplementary Fig. S4A). Pfetin expression was not observed in the 6 liver metastases examined or in primary high-risk GISTs that developed metastases between 13 and 30 months postsurgery (Supplementary Fig. S4B; Supplementary Table S1).

We used immunohistochemistry to evaluate pfetin expression *in situ*. Positive cells were diffusely stained with antipfetin antibody in membrane and cytoplasm. All nine G-GISTs expressed pfetin, whereas none of the eight P-GISTs did (Fig. 2B). Pfetin expression was not observed in neighboring host cells or in the intestinal cells of Cajal (Supplementary Fig. S5).

Real-time RT-PCR revealed pfetin mRNA levels were higher in G-GISTs than in P-GISTs (Fig. 2C). However, the difference between the P- and G- GIST group was less obvious at the mRNA than at the protein level, suggesting that pfetin expression is partially regulated at the mRNA level, and that

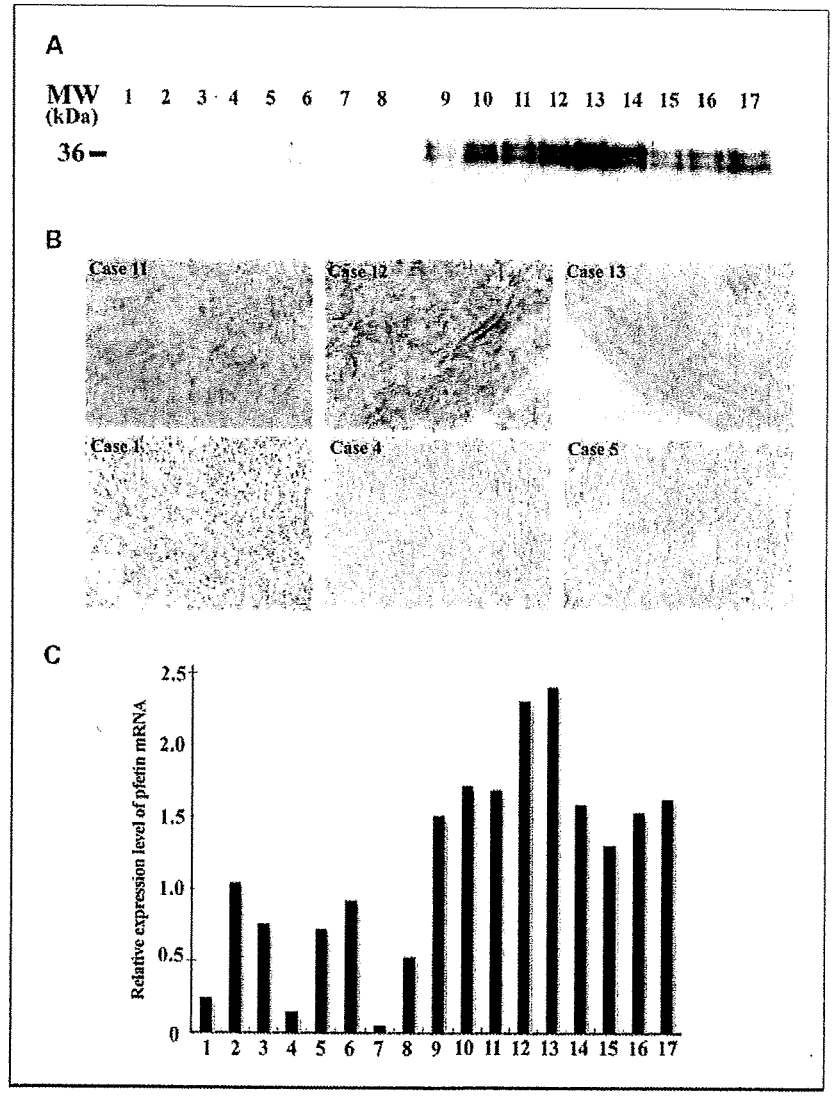

Fig. 2. Validation of the differential expression of pfetin. G-GISTs expressed pfetin at significantly higher levels than P-GISTs. *A*, Western blotting. Case numbers correspond to those in Fig. 1. *B*, immunohistochemistry; pfetin is overexpressed in G-GISTs (*top*), whereas it is not expressed in P-GISTs (*bottom*). *C*, the pfetin mRNA expression levels detected in the 17 GIST samples examined. P-GIST, 1:8; G-GIST, 9:17.

posttranscriptional regulation may also play an important role in pfetin expression.

As pfetin expression has been reported to correlate with c-kit mutation status (33), we examined 39 primary GISTs for the presence of c-kit and PDGFRA gene mutations and monitored their pfetin expression levels by Western blotting (Supplementary Tables S1 and S3). Overexpression of pfetin was observed in 12 of 29 c-kit mutation positive cases and in 6 of 10 negative cases ($P = 0.389$; Supplementary Table S3). PDGFRA mutations were not detected in the series. We observed no significant correlation between pfetin expression and c-kit or PDGFR mutation status.

The immunohistochemical study of 210 GISTs revealed a strong correlation between pfetin expression and a number of clinicopathologic variables including the tumor size, mitotic index, MIB-1 index, degree of differentiation, and risk classification ($P < 0.0001$; Table 3). Moreover, distant metastasis was observed in a significantly higher proportion of patients with pfetin-negative tumors compared with those with pfetin-positive tumors (24 of 39 versus 12 of 171 cases; $P < 0.0001$), with a median follow-up period of 73 months. The 5-year metastasis-free survival rate was significantly higher in the pfetin-positive than in the negative group overall (93.9% versus 36.2%; $P < 0.0001$; Fig. 3A; Table 3) as well as within each risk group (Fig. 3B-D). Multivariate analysis revealed that pfetin expression was a powerful predictor of disease-specific survival (Table 3). Note that high-risk cases were divided into two groups, the pfetin-positive and the pfetin-negative group, the latter having a worse prognosis. Furthermore, tumor-specific survival was statistically significantly longer in the pfetin-positive compared with the pfetin-negative group ($P < 0.0001$; Table 3; Supplementary Fig. S6). These data

Table 3. Univariate and multivariate analysis of prognostic factor and the relationship between clinicopathologic variables and pfetin expression

Variable	Number of cases	Metastasis-free survival		Tumor-specific survival		Multivariate analysis of metastasis free survival by Cox regression			Pfetin positive (no. cases)	Pfetin negative (no. cases)	Correlation (pfetin) χ^2 (*P*)
		5 y (%)	Log-rank (*P*)	5-y (%)	Log-rank (*P*)	*P*	Relative risk	95% CI			
Age			0.3290		0.8350						0.5220
<60	112	85.8 ± 3.7		91.8 ± 3.0					93	19	
60<	98	80.7 ± 4.2		95.6 ± 2.2					78	20	
Sex			0.4420		0.0393						0.8909
F	99	86.5 ± 3.7		96.1 ± 2.2					81	18	
M	111	80.4 ± 4.2		91.6 ± 2.8					90	21	
Site			0.0001		0.1655						0.4776
Stomach	170	87.8 ± 2.7		93.9 ± 2.0					140	30	
Nonstomach	40	60.7 ± 9.2		92.2 ± 5.3		0.0270	2.21	1.09-4.49	31	9	
Histology			0.1003		0.2068						0.5153
Spindle	189	85.4 ± 2.8		93.6 ± 2.0					155	34	
Epithelioid	21	67.7 ± 10.9		95.2 ± 4.6					16	5	
Size			<0.0001		<0.0001						<0.0001
<5 cm	128	92.4 ± 2.6		98.1 ± 1.3					112	16	
5-10 cm	63	76.6 ± 0.4		93.7 ± 3.5					51	12	
15 cm<	19	45.0 ± 11.9		60.5 ± 13.0		0.0070	2.05	1.22-3.44	8	11	
Necrosis			<0.0001		0.0034						0.0070
+	19	43.0 ± 12.7		76.9 ± 11.7					10	30	
-	191	86.9 ± 2.7		95.1 ± 1.7					161	9	
Miosis			<0.0001		<0.0001						<0.0001
<5/50HPF*	148	95.5 ± 2.0		98.2 ± 1.3					136	12	
5-10/50HPF	33	80.4 ± 7.2		96.8 ± 3.2					26	7	
5/50HPF<	29	29.9 ± 9.0		68.2 ± 9.4					9	20	
MIB-1			<0.0001		<0.0001						<0.0001
<9%	164	96.0 ± 1.8		98.4 ± 1.1					152	12	
10-29%	19	51.4 ± 12.6		85.6 ± 9.7					11	8	
<30%	27	32.5 ± 9.6		70.8 ± 9.4					8	19	
Differentiation			<0.0001		<0.0001						<0.0001
Score 1	161	95.9 ± 1.8		98.4 ± 1.1					149	12	
Score 2	49	43.5 ± 7.7		78.1 ± 6.6		<0.0001	10.40	3.68-29.45	22	27	
Risk classification			<0.0001		<0.0001						<0.0001
Low	110	97.7 ± 1.6		98.9 ± 1.1					100	10	
Intermediate	46	90.6 ± 5.2		97.0 ± 3.0					44	2	
High	54	48.5 ± 7.4		80.6 ± 6.0					27	27	
Pfetin			<0.0001		<0.0001						
Positive	171	93.9 ± 2.0		97.2 ± 1.4							
Negative	39	36.2 ± 8.7		76.5 ± 7.9		0.0020	3.75	1.60-8.81			

Abbreviation: 95% CI, 95% confidence interval.
*HPF, high-power field.