Prediction of Sensitivity to STI571 among Chronic Myeloid Leukemia Patients by Genome-wide cDNA Microarray Analysis

Yasuyuki Kaneta,[1] Yoshitoyo Kagami,[2, 12] Toyomasa Katagiri,[1, 12] Tatsuhiko Tsunoda,[3] Itsuro Jin-nai,[4] Hirokuni Taguchi,[5] Hisamaru Hirai,[6] Kazunori Ohnishi,[7] Takanori Ueda,[8] Nobuhiko Emi,[9] Akihiro Tomida,[10] Takashi Tsuruo,[10] Yusuke Nakamura[1, 13] and Ryuzo Ohno[11]

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, [2]Department of Hematology and Chemotherapy, Aichi Cancer Center Hospital, Kanokoden, Chikusa-ku, Nagoya 464-8681, [3]Laboratory of Medical Informatics, SNP Research Center, RIKEN (Institute of Physical and Chemical Research), 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, [4]Department of Hematology, Atomic Bomb Disease Institute, Nagasaki University School of Medicine, 1-12-4 Sakamoto, Nagasaki 852-8523, [5]Third Department of Internal Medicine, Kochi Medical School, 185-1 Kohasu, Okoh-cho, Nankoku, Kochi 783-8505, [6]Third Department of Internal Medicine, Faculty of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, [7]Department of Internal Medicine III, Hamamatsu University School of Medicine, 3600 Handa-cho, Hamamatsu 431-3192, [8]First Department of Internal Medicine, Fukui Medical School, 23-3 Shimoaizuki, Matsuoka-cho, Yoshida-gun, Fukui 910-1104, [9]The First Department of Internal Medicine, Nagoya University School of Medicine, 65 Tsurumai, Showa-ku, Nagoya 466-0064, [10]Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032 and [11]Aichi Cancer Center Hospital, Kanokoden, Chikusa-ku, Nagoya 464-8681

One of the most critical issues to be solved in regard to cancer chemotherapy is the establishment of ways to predict the efficacy of anti-cancer drugs for individual patients. To develop a prediction system based on expression of specific genes, we analyzed expression profiles of mononuclear cells from 18 chronic myeloid leukemia (CML) patients who were treated with the tyrosine kinase inhibitor STI571. cDNA microarrays representing 23 040 genes identified 79 genes that were expressed differentially between responders and non-responders to STI571. On the basis of the expression patterns of 15 or 30 of these genes among the patients, we developed a "Prediction Score" system that could clearly separate the responder group from the non-responder group. Verification of this system using four additional ("test") cases succeeded in predicting the response of each of those four patients to the drug. These results provide the first evidence that gene-expression profiles can predict sensitivity of CML cells to STI571, and may eventually lead to the achievement of "personalized therapy" for this disease.

Key words: cDNA microarray — Chronic myeloid leukemia — STI571 — Prediction score — Chemosensitivity

Chronic myeloid leukemia (CML) is a clonal disorder arising from neoplastic transformation of hematopoietic stem cells, most of which are characterized by the presence of a Philadelphia chromosome (Ph) and by constitutive activation of BCR-ABL tyrosine kinase.[1] CML progresses through three phases; chronic phase, accelerated phase and invariably fetal blast crisis. Conventional therapeutic options include interferon-α and allogenic stem-cell transplantation (SCT). Interferon-α prolongs overall survival, but has considerable adverse effects. SCT is the only curative treatment, but is associated with substantial morbidity and is limited to patients with suitable donors. Thus, the prognosis of CML is still poor.

Development of the ABL-selective tyrosine kinase inhibitor STI571 (imanitib; Glivec; Novartis Pharmaceuticals, Basel, Switzerland) was an important advance in the management of CML.[2, 3] With this drug, around 90% of CML patients are induced into hematological complete remission, and in more than 60% of patients Ph chromosome-positive leukemia cells are completely or partially reduced without severe adverse effects.[4] Thus, STI571 has become the first choice drug for the treatment of CML, and its promising effects make it difficult to decide the timing for SCT.[5] Moreover, time and medical cost are wasted if the drug is ineffective, and non-responsive patients risk losing a chance for alternative chemotherapy. Therefore, accurate prediction regarding effectiveness of a specific therapy is of critical importance for CML patients. Recent studies have demonstrated that information generated by cDNA microarray analysis of gene expression in

human tumors can provide more accurate data as to the nature of cancer cells than traditional histopathological methods are able to supply.[6-8] The promise of such information lies in its potential for improving clinical strategies for treating neoplastic diseases.

With this in mind we applied a microarray of human cDNAs consisting of 23 040 transcribed elements to analyze gene-expression profiles in CML cells, with a view to developing a novel system for predicting responsiveness of an individual CML patient to treatment with STI571. We identified a group of genes as differentially expressed among 12 patients belonging to the "good responder" category and six who showed no response to the drug, and established a "Prediction Score" system that correctly predicted the responsiveness or non-responsiveness among four additional test cases. Our results suggest that the expression levels of a set of genes selected in this way can determine the fate of CML exposed to STI571, and that such information may lead to "personalized therapy" and thereby improve the quality of life and prognosis of CML patients.

We obtained peripheral blood samples with informed consent from 22 Japanese adult CML patients prior to treatment with STI571. Each patient was then enrolled into a phase II study of STI571. mRNA from eighteen samples in which more than 65% of cells had been posi-

tive for the Ph chromosome prior to treatment, by means of a FISH analysis detecting a *bcr/abl* fusion gene,[9] were analyzed on our cDNA-microarray system. Fabrication of our cDNA-microarray system containing 23 040 cDNAs
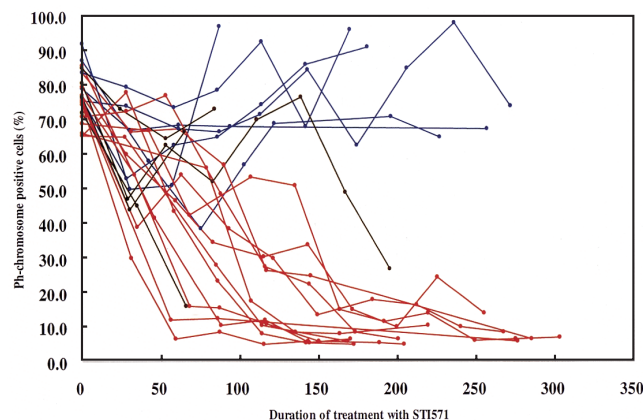

Fig. 1. Cytogenetic responses to STI571 treatment among 22 CML patients. Each line represents the response of an individual patient; blue lines indicate non-responders and red lines indicate responders. Black lines indicate test cases.

Table I. Clinicopathological Features of Patients Examined

Patient's ID	Age	Sex	Response	Prediction	Phase
CML003	66	M	Responder	Learning	Chronic
CML004	55	F	Responder	Learning	Chronic
CML008	61	F	Responder	Learning	Chronic
CML009	68	M	Responder	Test	Chronic
CML010	56	M	Responder	Learning	Chronic
CML013	59	F	Non-responder	Learning	Chronic
CML014	47	M	Responder	Learning	Chronic
CML015	63	F	Responder	Test	Chronic
CML018	57	M	Non-responder	Learning	Chronic
CML019	23	M	Non-responder	Learning	Chronic
CML021	57	M	Responder	Learning	Chronic
CML025	44	M	Non-responder	Learning	Chronic
CML027	35	M	Non-responder	Learning	Chronic
CML030	61	M	Responder	Learning	Chronic
CML033	56	M	Responder	Learning	Chronic
CML036	48	M	Responder	Learning	Chronic
CML047	32	F	Responder	Learning	Chronic
CML050	38	M	Non-responder	Learning	Blast crisis
CML054	32	M	Responder	Learning	Chronic
CML056	46	F	Responder	Learning	Blast crisis
CML080	59	F	Non-responder	Test	Accelerated
CML197	30	F	Non-responder	Test	Accelerated

Response, response to STI571 treatment; Learning, samples used to develop the prediction system; Test, samples used for test cases.
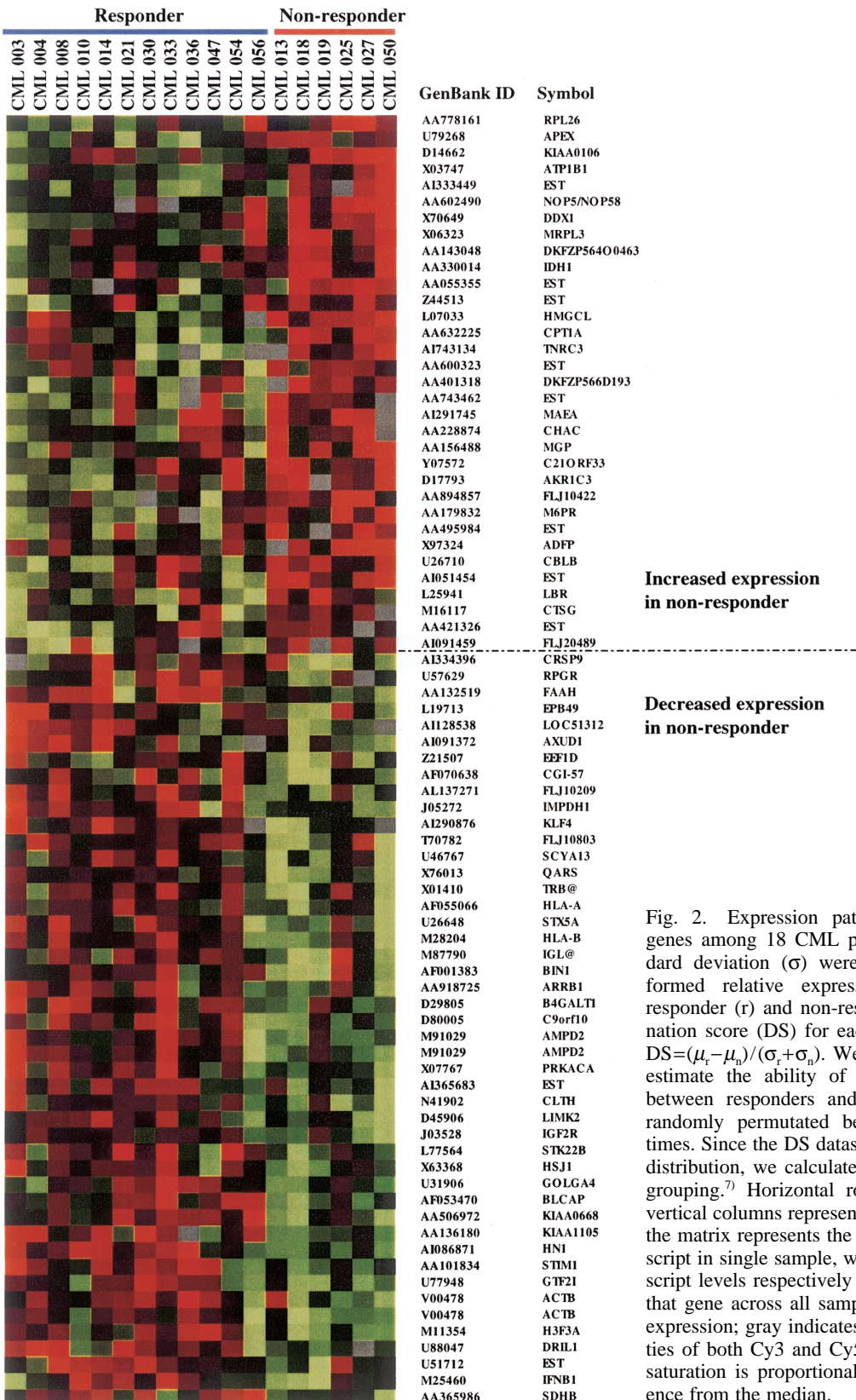
Responder — Non-responder

CML 003, CML 004, CML 008, CML 010, CML 014, CML 021, CML 030, CML 033, CML 036, CML 047, CML 054, CML 056, CML 013, CML 018, CML 019, CML 025, CML 027, CML 050

GenBank ID	Symbol
AA778161	RPL26
U79268	APEX
D14662	KIAA0106
X03747	ATP1B1
AI333449	EST
AA602490	NOP5/NOP58
X70649	DDX1
X06323	MRPL3
AA143048	DKFZP564O0463
AA330014	IDH1
AA055355	EST
Z44513	EST
L07033	HMGCL
AA632225	CPT1A
AI743134	TNRC3
AA600323	EST
AA401318	DKFZP566D193
AA743462	EST
AI291745	MAEA
AA228874	CHAC
AA156488	MGP
Y07572	C21ORF33
D17793	AKR1C3
AA894857	FLJ10422
AA179832	M6PR
AA495984	EST
X97324	ADFP
U26710	CBLB
AI051454	EST
L25941	LBR
M16117	CTSG
AA421326	EST
AI091459	FLJ20489

Increased expression in non-responder

- -

Decreased expression in non-responder

GenBank ID	Symbol
AI334396	CRSP9
U57629	RPGR
AA132519	FAAH
L19713	EPB49
AI128538	LOC51312
AI091372	AXUD1
Z21507	EEF1D
AF070638	CGI-57
AL137271	FLJ10209
J05272	IMPDH1
AI290876	KLF4
T70782	FLJ10803
U46767	SCYA13
X76013	QARS
X01410	TRB@
AF055066	HLA-A
U26648	STX5A
M28204	HLA-B
M87790	IGL@
AF001383	BIN1
AA918725	ARRB1
D29805	B4GALT1
D80005	C9orf10
M91029	AMPD2
M91029	AMPD2
X07767	PRKACA
AI365683	EST
N41902	CLTH
D45906	LIMK2
J03528	IGF2R
L77564	STK22B
X63368	HSJ1
U31906	GOLGA4
AF053470	BLCAP
AA506972	KIAA0668
AA136180	KIAA1105
AI086871	HN1
AA101834	STIM1
U77948	GTF2I
V00478	ACTB
V00478	ACTB
M11354	H3F3A
U88047	DRIL1
U51712	EST
M25460	IFNB1
AA365986	SDHB

Fig. 2. Expression patterns of the 79 discriminating genes among 18 CML patients. The mean (μ) and standard deviation (σ) were calculated from the log-transformed relative expression ratios of each gene in responder (r) and non-responder (n) patients. A discrimination score (DS) for each gene was defined as follows: $DS = (\mu_r - \mu_n)/(\sigma_r + \sigma_n)$. We carried out permutation tests to estimate the ability of individual genes to distinguish between responders and non-responders; samples were randomly permutated between the two classes 10 000 times. Since the DS dataset of each gene showed a normal distribution, we calculated a P value for the user-defined grouping.[7] Horizontal rows represent individual genes; vertical columns represent individual samples. Each cell in the matrix represents the expression level of a single transcript in single sample, with red and green indicating transcript levels respectively above and below the median for that gene across all samples. Black represents unchanged expression; gray indicates no or slight expression (intensities of both Cy3 and Cy5 under the cut-off value). Color saturation is proportional to the magnitude of the difference from the median.

851

Table II. A List of 79 Candidate Genes for Development of the Prediction System

Rank	Permutation *P*-value	GenBank ID	Symbol	Gene name
1	0.0003	AI086871	HN1	Humanin
2	0.0003	D17793	AKR1C3	aldo-keto reductase family 1, member C3
3	0.0013	X76013	QARS	glutaminyl-tRNA synthetase
4	0.0015	AA136180	KIAA1105	KIAA1105 protein
5	0.0019	AA506972	KIAA0668	KIAA0668 protein
6	0.0020	AF053470	BLCAP	bladder cancer associated protein
7	0.0021	X97324	ADFP	adipose differentiation-related protein
8	0.0029	AA894857	FLJ10422	hypothetical protein FLJ10422
9	0.0038	L07033	HMGCL	3-hydroxymethyl-3-methylglutaryl-coenzyme A lyase
10	0.0040	AI051454	EST	EST
11	0.0063	AI290876	KLF4	Kruppel-like factor 4
12	0.0083	M11354	H3F3A	H3 histone, family 3A
13	0.0094	V00478	ACTB	actin, beta
14	0.0094	AA401318	DKFZP566D193	DKFZP566D193 protein
15	0.0101	U79268	APEX	APEX nuclease
16	0.0107	U88047	DRIL1	dead ringer (*Drosophila*)-like 1
17	0.0113	AF001383	BIN1	bridging integrator 1
18	0.0114	AA495984	EST	EST
19	0.0123	N41902	CLTH	Clathrin assembly lymphoid-myeloid leukemia gene
20	0.0127	AA179832	M6PR	mannose-6-phosphate receptor
21	0.0133	D14662	KIAA0106	anti-oxidant protein 2
22	0.0139	J03528	IGF2R	insulin-like growth factor 2 receptor
23	0.0151	AA330014	IDH1	isocitrate dehydrogenase 1 (NADP+), soluble
24	0.0154	AI333449	EST	EST
25	0.0156	AA365986	SDHB	succinate dehydrogenase complex, subunit B
26	0.0165	AI743134	TNRC3	trinucleotide repeat containing 3
27	0.0171	AA156488	MGP	KIAA1008 protein
28	0.0178	U26710	CBLB	Cas-Br-M ectropic retroviral transforming sequence b
29	0.0187	AA055355	EST	EST
30	0.0191	T70782	FLJ10803	hypothetical protein FLJ10803
31	0.0193	J05272	IMPDH1	IMP (inosine monophosphate) dehydrogenase 1
32	0.0197	AI091459	FLJ20489	hypothetical protein FLJ20489
33	0.0200	U77948	GTF2I	major histocompatibility complex, class I, B
42	0.0271	U31906	GOLGA4	golgi autoantigen, golgin subfamily a, 4
43	0.0272	AA743462	EST	EST
44	0.0279	U46767	SCYA13	small inducible cytokine subfamily A, member 13
45	0.0281	D29805	B4GALT1	beta 1,4-galactosyltransferase, polypeptide 1
46	0.0290	AA143048	DKFZP564O0463	DKFZP564O0463 protein
47	0.0299	V00478	ACTB	actin, beta
48	0.0314	X63368	HSJ1	heat shock protein, neuronal DNAJ-like 1
49	0.0315	X06323	MRPL3	mitochondrial ribosomal protein L3
50	0.0320	D80005	C9orf10	C9orf10 protein
51	0.0327	X70649	DDX1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1
52	0.0350	AA421326	EST	*Homo sapiens* cDNA: FLJ21918 fis, clone HEP04006
53	0.0353	AF055066	HLA-A	major histocompatibility complex, class I, A
54	0.0359	AA101834	STIM1	stromal interaction molecule 1
55	0.0360	M91029	AMPD2	adenosine monophosphate deaminase 2
56	0.0361	U26648	STX5A	syntaxin 5A
57	0.0366	M25460	IFNB1	interferon, beta 1, fibroblast
58	0.0370	AI291745	MAEA	macrophage erythroblast attacher
59	0.0372	L25941	LBR	lamin B receptor
60	0.0373	D45906	LIMK2	LIM domain kinase 2
61	0.0387	AI365683	EST	*Homo sapiens* PAC clone RP4-751H13 from 7q35-qter

Table II. (Continued)

Rank	Permutation P-value	GenBank ID	Symbol	Gene name
62	0.0391	AA778161	RPL26	ribosomal protein L26
63	0.0395	AL137271	FLJ10209	hypothetical protein FLJ10209
64	0.0407	AA132519	FAAH	EST
65	0.0415	Y07572	C21ORF33	ES1 (zebrafish) protein, human homolog of
66	0.0427	Z44513	EST	EST
67	0.0432	X07767	PRKACA	protein kinase, cAMP-dependent, catalytic, alpha
68	0.0433	L19713	EPB49	erythrocyte membrane protein band 4.9 (dematin)
69	0.0439	M91029	AMPD2	adenosine monophosphate deaminase 2
70	0.0439	U51712	EST	EST
71	0.0442	AI334396	CRSP9	cofactor required for Sp1 transcriptional activation
72	0.0442	AA600323	EST	EST
73	0.0442	L77564	STK22B	serine/threonine kinase 22B
74	0.0444	X01410	TRB@	T cell receptor beta locus
75	0.0446	Z21507	EEF1D	eukaryotic translation elongation factor 1 delta
76	0.0446	U57629	RPGR	retinitis pigmentosa GTPase regulator
77	0.0454	AA918725	ARRB1	arrestin, beta 1
78	0.0458	AA602490	NOP5/NOP58	nucleolar protein NOP5/NOP58
79	0.0461	M87790	IGL@	immunoglobulin lambda locus

Information was retrieved from Unigene database in National Center for Biotechnology Information (NCBI) (build#131).

was described previously.[10] We prepared mononuclear cells using Ficoll (Amersham Biosciences, Buckinghamshire, UK) and extracted total RNA using TRIzol (Life Technologies, Inc., Grand Island, NY) according to the manufacturer's instructions. After treatment with DNase I (Nippon Gene, Tokyo), T7-based RNA amplification was carried out as described previously.[11] Two rounds of amplification using 2 μg of total RNA as starting material yielded 40–100 μg of amplified RNA (aRNA). For control samples we also performed two rounds of T7-based RNA amplification to obtain sufficient amounts of aRNA. RNA amplified by this method accurately reflects the proportions in the original RNA source, as we had confirmed earlier by semi-quantitative RT-PCR experiments, in which data from microarrays were consistent with results from RT-PCR whether total RNA or aRNA was used as the template.[11] Labeling, hybridization, washing, scanning, and quantification of signals were performed as described previously[11] except that all processes were carried out with an Automated Slide Processor.[10] Sixteen patients with CML in the chronic phase were treated with 400 mg/day of STI571 and two patients in blast crisis were treated with 600 mg/day. We determined the clinical response to STI571 by cytogenetic criteria; that is, by the percentage of peripheral blood cells positive for Ph chromosome by the FISH analysis.[4] The 12 patients who showed major cytogenetic responses (less than 35% of cells remaining positive for the Ph chromosome) were classified as responders (red lines in Fig. 1), whereas the six patients with more than 65% of cells still positive for the Ph chromosome after 5 months of STI571 treatment were considered non-responders (blue lines in Fig. 1). The remaining four were reserved to test the predictive scoring system later (black lines in Fig. 1). Of the 22, two "learning" cases were in blast crisis phase and two "test" cases were in accelerated phase (Table I), and their cytogenetic responses were analyzed within 12 weeks after the start of treatment, because STI571 was clinically ineffective and was discontinued within 12 weeks (Table I, Fig. 1). As controls we used a mixture of mononuclear cells from peripheral blood of 11 healthy volunteers.

We first selected genes using two criteria; (i) signal intensities higher than the cut-off level in at least 80% of the cases; (ii) $|Med_r - Med_n| \geq 0.5$, where Med indicates the median derived from log-transformed relative expression ratios in responders or non-responders. Then we carried out a permutation test to select genes that were useful for separation of the responder group from the non-responder group (see the legend of Fig. 2). As a result, 79 genes were listed as candidates that showed a permutation P-value of less than 0.05. Expression levels were increased for 33 of those genes and decreased for the other 46 in the non-responder group, as compared to the responder group (Fig. 2).

Using this information we attempted to establish a scoring system to predict the efficacy of STI571 treatment. We calculated the prediction score according to procedures described previously.[7, 12] Each gene (g_i) votes for either
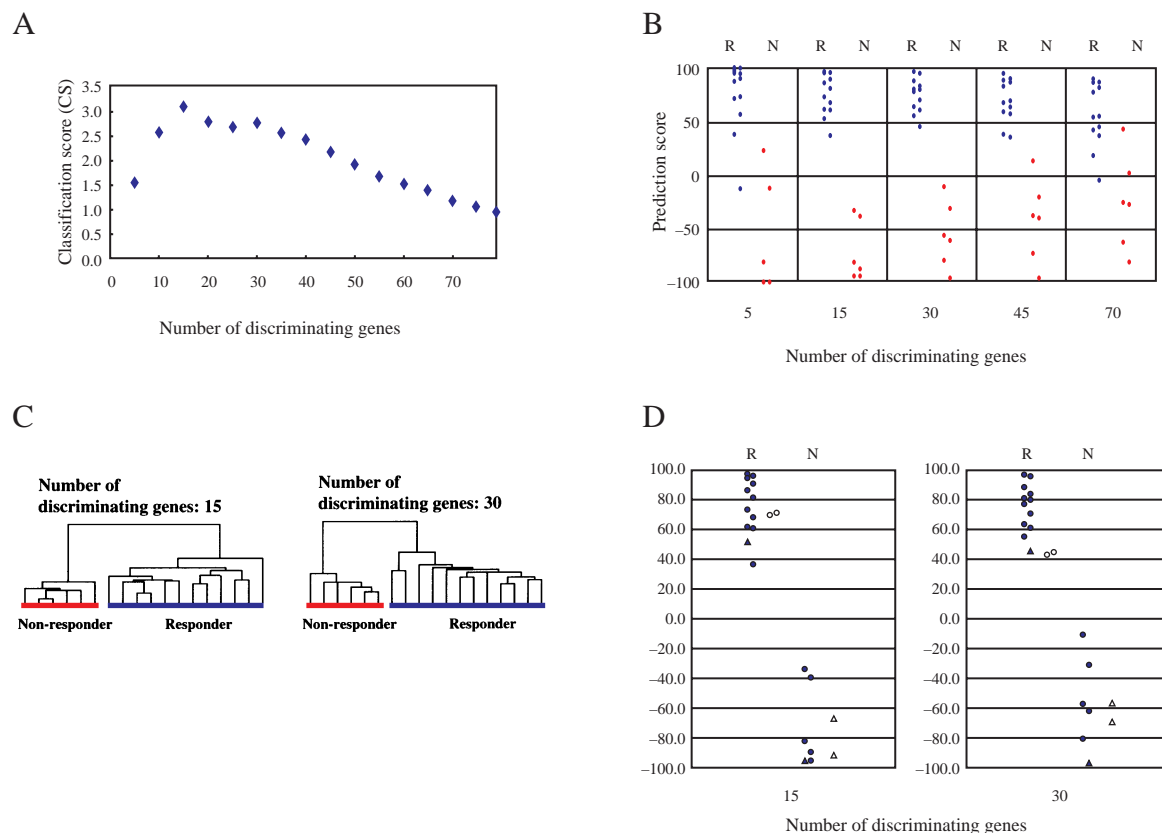
Fig. 3. (A) Optimization of the number of discriminating genes. The classification score (CS) was calculated by using the prediction score of responders (PS$_r$) and non-responders (PS$_n$) in each gene set, as follows: CS$=(\mu_{PSr}-\mu_{PSn})/(\sigma_{PSr}+\sigma_{PSn})$. A larger value of CS indicates better separation of the two groups by the predictive-scoring system. (B) Different prediction scores appear when the number of discriminating genes is changed. R, responder; N, non-responder. (C) Cluster analysis of the predictive 15- or 30-gene sets. All samples were classified according to their sensitivity to STI571. (D) Prediction scores for individual patients. Filled circles and filled triangles indicate scores in cross-validation cases of patients whose expression data were used for selecting discriminating genes (learning). Open circles and open triangles represent scores for four additional (test) cases. Circles indicate CML patients in chronic phase and triangles show CML patients in blast crisis (learning) and accelerated phases (test), respectively. High absolute values indicate high confidence.

responder or non-responder depending on whether the expression level (x_i) in the sample is closer to the mean expression level of responders or non-responders in reference samples. The magnitude of the vote (v_i) reflects the deviation of the expression level in the sample from the average of the two classes: $V_i=|x_i-(\mu_r+\mu_n)/2|$. We summed the votes to obtain total votes for the responder (V_r) and non-responder (V_n), and calculated PS values as follows: PS$=((V_r-V_n)/(V_r+V_n))\times100$, reflecting the margin of victory in the direction of either responder or non-responder. PS values range from -100 to 100; a higher absolute value of PS reflects a stronger prediction. Next we rank-ordered the 79 candidate genes on the basis of the magnitude of their permutation P-values (Table II) and calculated the prediction score by the leave-one-out test for cross-validation using the top 5, 10, 15, 20, 25, 30, 35,

40, 45, 50, 55, 60, 65, 70, 75, and 79 genes on the rank-ordered list. For the leave-one-out test, one sample is withheld, the permutation P-value and mean expression levels are calculated using the remaining samples, and the class of the withheld sample is subsequently evaluated by calculating its prediction score. We repeated this procedure for each of the 18 samples. Then, to determine the number of discriminating genes that provided the best separation of the two groups, we calculated a classification score (CS) for each gene set (see the legend of Fig. 3A).

The number of genes used for calculation influenced the power for separation of the two groups. We obtained the best separation when we used the top 15 or 30 genes in our candidate list for calculation of the scores (Fig. 3A). The "Prediction Score" system using these two sets of genes clearly separated the two patient groups (Fig. 3B).

Hierarchical clustering using the same gene sets was also able to classify the groups with regard to STI571 sensitivity (Fig. 3C). This analysis was performed using web-available software ("cluster" and "treeview") written by M. Eisen (http://genome-www5/stanford.edu/MicroArray/SMD/restech.html). Before the clustering algorithm was applied, the fluorescence ratio for each spot was first log-transformed and then the data for each sample were median-centered to remove experimental biases.

To validate this prediction system, we investigated four additional ("test") cases that were completely independent of the 18 "learning" cases used for establishing the system. We examined gene-expression profiles in each of these four blood samples and then calculated a prediction score for each of them using the panels of 15 or 30 discriminating genes. As shown in Fig. 3D, responsiveness of each of these four patients to STI571 was predicted accurately.

Treatment of CML patients with STI571 provides a considerable advantage over treatment with interferon-α, because the degree of cytogenetic response obtained by STI571 is clearly higher, with less severe adverse effects.[13] However, as around 40% of CML patients fail to achieve major cytogenetic response,[4] it is important to establish a way to predict the efficacy of STI571 before therapy is undertaken for an individual patient.

In this study we profiled the gene expression patterns of mononuclear cells from CML patients using a comprehensive cDNA-microarray system containing 23 040 genes, with a view to establishing a "Prediction Score" system. We identified 79 genes that were differentially expressed among patients who showed good response as opposed to poor response to STI571, and ranked them by the permutation test (P-values of <0.05). Then we attempted to optimize the power to separate the two groups by selecting a subset of the discriminating genes on the basis of CS. We obtained the best CS using subsets of 15 or 30 discriminating genes; the scoring system based on these two subsets separated the two groups very clearly. Our scoring system was also able to predict accurately the response to STI571 of four additional cases. Although four patients who were in accelerated and blast crisis phases of CML were included, our scoring system classified all cases, as to their chemosensitivity to STI571. We believe that the use of CS is a reasonable approach to selecting appropriate indicators for predictive scores.

The 79 genes that showed different expression levels in responders versus non-responders might provide interesting insight into the biological mechanism underlying the response to STI571 in CML. Among these genes, Apex nuclease (APEX) was over-expressed in the non-responder group. APEX is the major apurinic/apyrimidinic endonuclease (Ap endo), with a key function in the DNA-repair system that confers resistance to ionizing radiation and alkylating agents in human cell lines.[14] APEX, which stimulates the binding of MYB (v-myb myeloblastosis viral oncogene homolog) to DNA, is a known accelerator of proliferation and was up-regulated in all CML cells in our experiments.[15]

On the other hand, Kruppel-like factor 4 (KLF4) was suppressed in the non-responder group. This gene is a zinc finger-containing transcription factor, enriched in epithelial cells, that is known to suppress cell proliferation.[16] Thus, down-regulated expression of KLF4 in the non-responders might abrogate regulation of the cell cycle. We suggest that the 79 genes that may affect sensitivity of CML cells to STI571 could serve as molecular targets for overcoming chemoresistance and also for development of novel drugs.

Although adverse drug reactions caused by STI571 are much less severe than with other anti-cancer drugs, its long-term use may ruin the chance of a non-responder to benefit from alternative therapies. Hence, we believe that our prediction system should provide an opportunity for potential non-responders to achieve a better prognosis and a better quality of life, although certainly a larger-scale study is warranted. Our data suggest, however, that the goal of "personalized medicine," giving the right drug to each patient, may be achievable by selecting a set of genes for its predictive value according to the approach shown here. However, to establish optimal prediction we should consider including genotypic information such as single nucleotide polymorphisms (SNPs) that might influence the metabolism of the drug in question.

REFERENCES

1) Faderl, S., Talpaz, M., Estrov, Z., O'Brien, S., Kurzrock, R. and Kantarjian, H. M. The biology of chronic myeloid leukemia. *N. Engl. J. Med.*, **341**, 164–172 (1999).
2) Buchdunger, E., Matter, A. and Druker, B. J. Bcr-Abl inhi-

bition as a modality of CML therapeutics. *Biochim. Biophys. Acta*, **1551**, M11–M18 (2001).

3) Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J. and Lydon, N. B. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat. Med.*, **2**, 561–566 (1996).

4) Druker, B. J., Talpaz, M., Resta, D. J., Peng, B., Buchdunger, E., Ford, J. M., Lydon, N. B., Kantarjian, H., Capdeville, R., Ohno-Jones, S. and Sawyers, C. L. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.*, **344**, 1031–1037 (2001).

5) Goldman, J. M. and Druker, B. J. Chronic myeloid leukemia: current treatment options. *Blood*, **98**, 2039–2042 (2001).

6) Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R. and Korsmeyer, S. J. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat. Genet.*, **30**, 41–47 (2002).

7) Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537 (1999).

8) Hofmann, W. K., de Vos, S., Elashoff, D., Gschaidmeier, H., Hoelzer, D., Koeffler, H. P. and Ottmann, O. G. Relation between resistance of Philadelphia-chromosome-positive acute lymphoblastic leukaemia to the tyrosine kinase inhibitor STI571 and gene-expression profiles: a gene-expression study. *Lancet*, **359**, 481–486 (2002).

9) Yanagi, M., Shinjo, K., Takeshita, A., Tobita, T., Yano, K., Kobayashi, M., Terasaki, H., Naoe, T., Ohnishi, K. and Ohno, R. Simple and reliably sensitive diagnosis and monitoring of Philadelphia chromosome-positive cells in chronic myeloid leukemia by interphase fluorescence *in situ* hybridization of peripheral blood cells. *Leukemia*, **13**, 542–552 (1999).

10) Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y. and Nakamura, Y. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. *Cancer Res.*, **61**, 2129–2137 (2001).

11) Ono, K., Tanaka, T., Tsunoda, T., Kitahara, O., Kihara, C., Okamoto, A., Ochiai, K., Takagi, T. and Nakamura, Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Res.*, **60**, 5007–5011 (2000).

12) MacDonald, T. J., Brown, K. M., LaFleur, B., Peterson, K., Lawlor, C., Chen, Y., Packer, R. J., Cogen, P. and Stephan, D. A. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat. Genet.*, **29**, 143–152 (2001).

13) Goldman, J. M. and Melo, J. V. Targeting the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.*, **344**, 1084–1086 (2001).

14) Bobola, M. S., Blank, A., Berger, M. S., Stevens, B. A. and Silber, J. R. Apurinic/apyrimidinic endonuclease activity is elevated in human adult gliomas. *Clin. Cancer Res.*, **7**, 3510–3518 (2001).

15) Evans, A. R., Limp-Foster, M. and Kelley, M. R. Going APE over ref-1. *Mutat. Res.*, **461**, 83–108 (2000).

16) Dang, D. T., Mahatan, C. S., Dang, L. H., Agboola, I. A. and Yang, V. W. Expression of the gut-enriched Kruppel-like factor (Kruppel-like factor 4) gene in the human colon cancer cell line RKO is dependent on CDX2. *Oncogene*, **20**, 4884–4890 (2001).