

Genetic Alterations in the Tyrosine Kinase Transcriptome of Human Cancer Cell Lines

Jens E. Ruhe,¹ Sylvia Streit,¹ Stefan Hart,¹ Chee-Hong Wong,² Katja Specht,⁴ Piotr Knyazev,⁵ Tatjana Knyazeva,⁵ Liang Seah Tay,¹ Hooi Linn Loo,¹ Priscilla Foo,¹ Winnie Wong,¹ Sharon Pok,¹ Shu Jing Lim,¹ Huimin Ong,¹ Ming Luo,³ Han Kiat Ho,¹ Kaitian Peng,¹ Tze Chuen Lee,² Martin Bezler,⁵ Christian Mann,⁵ Silvia Gaertner,⁵ Heinz Hoefler,⁴ Stefano Iacobelli,⁶ Stephan Peter,⁷ Alice Tay,³ Sydney Brenner,^{1,3} Byrappa Venkatesh,³ and Axel Ullrich^{1,5}

¹Institute of Medical Biology, ²Bioinformatics Institute, and ³Institute of Molecular and Cell Biology, Singapore, Singapore;

⁴Department of Pathology, Technical University of Munich, Munich, Germany; ⁵Max-Planck-Institute for Biochemistry,

Martinsried, Germany; ⁶Department of Oncology, University of Chieti Medical School, Chieti, Italy;

and ⁷Department of Urology, Klinikum Darmstadt, Darmstadt, Germany

Abstract

Protein tyrosine kinases (PTKs) play a critical role in the manifestation of cancer cell properties, and respective signaling mechanisms have been studied extensively on immortalized tumor cells. To characterize and analyze commonly used cancer cell lines with regard to variations in the primary structure of all expressed PTKs, we conducted a cDNA-based sequence analysis of the entire tyrosine kinase transcriptome of 254 established tumor cell lines. The profiles of cell line intrinsic PTK transcript alterations and the evaluation of 155 identified polymorphisms and 234 somatic mutations are made available in a database designated “Tykiva” (*tyrosine kinome variant*). Tissue distribution analysis and/or the localization within defined protein domains indicate functional relevance of several genetic alterations. The cysteine replacement of the highly conserved Y367 residue in fibroblast growth factor receptor 4 or the Q26X nonsense mutation in the tumor-suppressor kinase CSK are examples, and may contribute to cell line-specific signaling characteristics and tumor progression. Moreover, known variants, such as epidermal growth factor receptor G719S, that were shown to mediate anticancer drug sensitivity could be detected in other than the previously reported tumor types. Our data therefore provide extensive system information for the design and interpretation of cell line-based cancer research, and may stimulate further investigations into broader clinical applications of current cancer therapeutics. [Cancer Res 2007;67(23):11368–76]

Introduction

For decades, research on established cancer cell lines has been the basis of pioneering discoveries in diverse areas of modern biology. Especially the fields of tumor biology and signal transduction research critically depended on studies of tumor-derived cell lines. Also, the development of currently used cancer

therapeutics, including protein tyrosine kinase (PTK) targeting agents, was made possible through the use of cancer cells in culture.

Because of the established relevance of PTKs for cancer initiation and progression, extensive screens for genomic sequence changes of the cancer kinome have been undertaken and yielded a diversity of sporadic alterations (1–8). Besides sporadic mutations that accumulate in the developing tumor due to cancer genome instability, DNA sequence polymorphisms have been linked to cancer aggressiveness and susceptibility (9). The fibroblast growth factor receptor 4 (FGFR4) G388R single nucleotide polymorphism (SNP), for instance, was found to be associated with increased cell motility *in vitro* as well as lymph node metastasis, advanced tumor stage, and reduced survival of breast cancer patients (10). Moreover, this SNP was shown to mediate resistance to adjuvant therapy in primary breast cancer, and its clinical relevance was also shown for soft tissue sarcoma, and head and neck, colon, and prostate cancers (11). Recent genome-wide association studies to identify SNPs connected with cancer susceptibility reflect the increasing recognition of polymorphisms as cancer-relevant parameters (12–14).

Through cDNA-based tyrosine kinase transcriptome (TKT) sequence evaluation, we performed a comprehensive characterization and analysis of widely used cancer cell lines with regard to variations in the primary structure of all expressed tyrosine kinases. For 254 established tumor cell lines of 19 tissue origins, we determined the profiles of both somatic mutations and germ line polymorphisms detectable in the transcripts of 90 PTK genes (15). The localization within the respective protein sequence, sequence comparisons with related genes, or the tissue distribution indicated potential functional relevance for individual genetic alterations that may define the cell line specific signaling characteristics and/or properties related to oncogenesis.

Materials and Methods

Samples. The panel of 254 cancer and 7 nontumorigenic cell lines was obtained from various sources (Supplementary Table S1). Poly(A)+ RNAs from diverse organs of 15 different cancer-free individuals were purchased from Ambion (Supplementary Table S1); genomic DNA derived from blood of 90 noncancer donors was from Coriell Institute for Medical Research. Samples of 55 primary invasive breast carcinomas were obtained from H. Hoefler and S. Iacobelli. Biopsies of 55 kidney tumors and 55 prostate cancer specimens were provided by S. Peter (Supplementary Table S2).

RNA isolation and cDNA synthesis. Total RNA was extracted from exponentially growing cells using guanidinium thiocyanate (16).

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

Current address for J.E. Ruhe: U3 Pharma AG, Martinsried, Germany.

J.E. Ruhe, S. Streit, and S. Hart contributed equally to this work.

Requests for reprints: Jens E. Ruhe, U3 Pharma AG, Bunsenstrasse 1, 82152 Martinsried, Germany. E-mail: ruhe@u3pharma.com.

©2007 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-07-2703

Subsequently, mRNA was isolated via oligo(dT) cellulose and reverse transcribed using oligo(dT) primers (17).

Genes, PCR, and sequencing. With purified cDNAs prepared from 261 cell lines and 15 control tissues as templates, overlapping fragments covering the entire coding region of all 58 receptor tyrosine kinase (RTK) and 32 cytoplasmic tyrosine kinase genes (15) were PCR amplified in at least two independent experiments. For each fragment, PCR conditions were specifically optimized using pooled cDNAs in a first step, and cDNAs from six individual cell lines in a second step. Thereby, the efficiencies of three proofreading polymerases (Novagen, Roche, Invitrogen) in combination with or without additives (DMSO, betaine, TMA oxalate), various primer pairs and different annealing temperatures were compared, and best parameters applied to the full panel of samples.

PCR products for all genes except *IRR*, *MUSK*, *FGR*, and *SRMS*, for which we obtained no amplicons, were subjected to direct sequencing of both forward and reverse strands. For selected alterations, homologous fragments from cDNAs of 165 primary tumors were PCR amplified using the same primer pairs. To further verify potential germ line origin of given variants, additional primers were designed to PCR amplify the respective region from genomic DNA of 90 blood samples from healthy donors.

Primers for amplification and sequencing of cDNA (Supplementary Table S3) and genomic DNA (Supplementary Table S4) were designed using the Primer3 program⁸ and refer to National Center for Biotechnology Information (NCBI)⁹ reference sequence files with accession numbers provided in Supplementary Table S5.

Data analysis. Sequence differences to the NCBI⁹ reference sequence were identified via manual inspection of aligned electropherograms assisted by the Mutation Surveyor software package. In addition to nonsynonymous genetic alterations, we detected numerous silent sequence variations that are not presented and further analyzed here. It is noteworthy that 63 sequence differences to the NCBI⁹ database occurred in all cDNA samples we analyzed and represent either other isoforms or variants that might actually be wild-type rather than genetic alterations (Supplementary Table S6).

Results and Discussion

Characterization of PTK transcripts in 254 tumor cell lines.

To comprehensively characterize widely used tumor cell lines with regard to nonsilent alterations in all expressed tyrosine kinase genes, we evaluated the sequence of the entire TKT of 254 established cancer cell lines (Supplementary Table S1). These cell lines were derived from 19 different tissue origins (Fig. 1A), controls included seven nontumorigenic cell lines and 15 tissues from different organs of healthy individuals.

For each cell line and control sample, whole-cell cDNA was prepared and used for the amplification and direct sequencing of the complete PTK coding region. Despite the lack of normal tissue counterparts for the established tumor cell lines, we attempted to define the identified PTK transcript variations as somatic or germ line sequence differences. We declared those sequence differences as germ line polymorphisms that were either detected in our 22 controls or were previously reported as hereditary variants in the NCBI SNP database,⁹ Ensemble Genome Browser,¹⁰ the UniProtKB/Swiss-Prot database,¹¹ the SNP500Cancer database,¹² or the literature. Besides zygosity, cell line-specific variant profiles thus indicate germ line or somatic origin of the individual TKT sequence

variations. Representative examples are displayed for the skin-derived tumor cell lines A-375, BOW-G, C-32, C-8161, and Colo-16 (Fig. 1B). The characterization of all cell lines can be found in Supplementary Table S7.

Based on these data, we determined the absolute number and distribution of TKT-linked somatic mutations and germ line polymorphisms within the entire tumor cell line panel. For polymorphisms, we observed a Gaussian-like distribution with an average of 12.3 sequence variations per cancer cell line. In contrast, somatic mutations were unevenly distributed (Fig. 1C). No somatic alterations were detected in the TKT of 119 cancer cell lines, consistent with kinome mutations entirely absent in subsets of recently screened breast cancer, lung carcinoid, and testicular germ-cell tumor samples (1–3). In contrast, high frequencies of 9 to 14 somatic mutations in the transcribed tyrosine kinomes of LNCaP, Jurkat, MeWo, MKN-1, HCT-15, and DLD-1 might reflect a mutator phenotype (1). They are in agreement with sequence data of 24 cancer genes in the NCI-60 cell line panel that also showed HCT-15 to be one of the most frequently mutated tumor cell lines (18). With intermediate mutation rates for the other tumor cell lines, our data indicate an accumulation of somatic mutations in PTK transcripts of various cancer cell lines which may contribute to the progression characteristics of certain cancers.

In the entire panel of cancer cell lines and controls, we identified 389 nonsynonymous genetic variations within 39.85 of 72.08 MB of PTK coding sequence that were amplifiable and represent the detectable TKT of all samples. As alternative to the allocation to cell line-specific variant profiles, we grouped these sequence differences by genes and PTK subfamilies. Each variation was thereby specified regarding the spectrum of affected cell lines as well as the zygosity status and the presumable somatic versus germ line origin. These data are shown for FGFR4 (Fig. 2) that will be discussed below. The full information for all transcript variants and PTK genes can be obtained from Supplementary Table S8. Sequence specifications of frame shift alterations and insertions are provided in Supplementary Table S9.

Tykiva database for cancer cell line TKT analysis. Additionally, all data on the identified PTK transcript variants are compiled in the database designated Tykiva.¹³ Transcript variants can be specifically retrieved for each of the 254 tumor cell lines, 19 tissue origins/tumor types, or any of the 90 PTK genes. Somatic or germ line origin is indicated, and other cell lines carrying the same variant are referred to. In graphical gene representations, the localization of all detected variants is displayed in the context of the reference amino acid sequence as well as predicted protein domain structures according to Swiss-Prot data. Optionally projectable to the major known isoforms, these illustrations cross-reference our data to variant information from the NCBI SNP database,⁹ the Ensemble Genome Browser,¹⁰ the Swiss-Prot¹¹ and the Genbank¹⁴ databases, the KinMutBase,¹⁵ the IDbases,¹⁶ and the literature. By that, they comprise the current knowledge of nonsilent genetic variations in PTK genes.

The expressed PTK variants may define cell line-specific signaling characteristics and cancer-related cell properties. In the following sections, we therefore address tissue distribution and

⁸ http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

⁹ <http://www.ncbi.nlm.nih.gov>

¹⁰ <http://www.ensembl.org>

¹¹ <http://ca.expasy.org>

¹² <http://snp500gov.nci.nih.gov>

¹³ <http://tykiva.bii.a-star.edu.sg>

¹⁴ <http://www.ncbi.nlm.nih.gov/Genbank>

¹⁵ <http://bioinf.uta.fi/KinMutBase>

¹⁶ http://bioinf.uta.fi/base_root

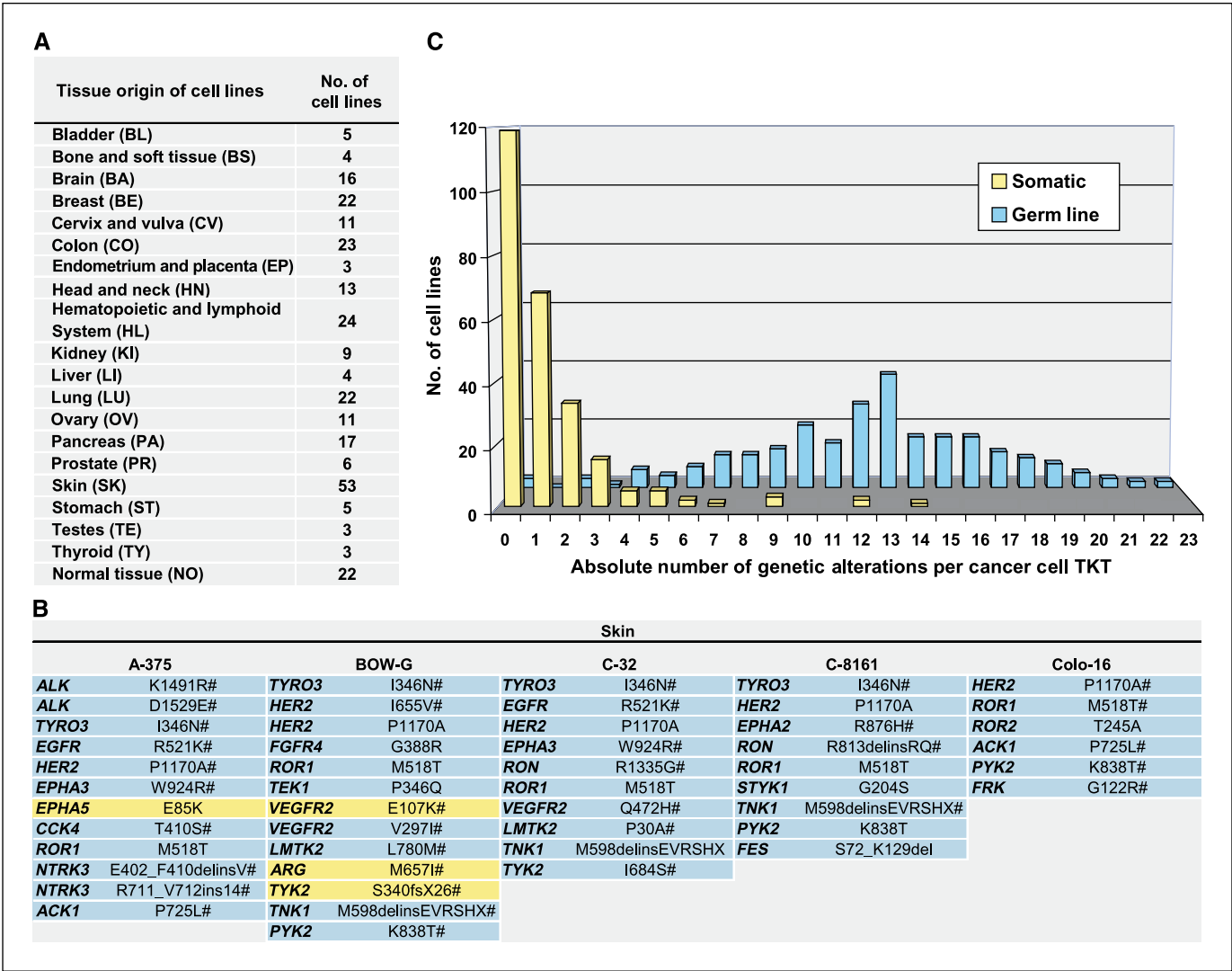


Figure 1. Characterization of tumor cell lines with regard to genetic alterations in the TKT. **A**, samples. The tissue origins and number of tumor cell lines derived thereof are summarized. **B**, patterns of genetic alterations. For each of the 254 tumor and seven control cell lines, the specific pattern of nonsynonymous genetic alterations within the TKT is provided (Supplementary Table S7) and exemplarily shown for five skin-derived tumor cell lines. Blue, germ line polymorphisms; yellow, somatic mutations. **C**, genetic alterations per TKT. The number of tumor cell lines with the indicated number of somatic (yellow) or germ line (blue) variations detected therein is illustrated.

localization of each polymorphism and somatic mutation within the respective protein sequence. Based on these data and the current literature, we discuss potential functional and/or clinical relevance for some of the identified genetic variants.

Characterization of 155 PTK gene sequence polymorphisms. According to our definition of somatic and germ line sequence variants, 155 of the 389 identified alterations were classified as sequence polymorphisms. They include 131 SNPs, 16 germ line deletions, and 8 insertions. Their overall frequencies and localization in distinct protein domains are summarized in Fig. 3A and B. Moreover, the occurrence frequency of each polymorphism in individual tumor types or control samples was determined. Occurrence frequency was thereby defined as the fraction of carriers of a given sequence variant and the number of cell lines with the same tissue origin that express the corresponding gene regardless of its genotype (paired numbers in Fig. 3C and Supplementary Table S10). It therefore reflects the expression

aspect of respective genes and alterations, as addressed by our cDNA analysis.

Of the 131 missense substitutions, 100 had been reported previously. However, only 12 of them, as well as two deletions, have been connected with cancer thus far (Supplementary Table S10). It is noteworthy that five of eight novel deletions involve entire exons (Supplementary Table S10) and most likely represent splice variants. Such variants could preferentially be detected because of the use of cDNA as sequencing target. Moreover, other transcription-related mechanisms such as epigenetic gene silencing or mRNA stability are also reflected by cDNA, and genetic alterations identified therein are thus likely to be expressed within the cell. However, a disadvantage of our approach is that it does not detect fusion kinase gene or amplified kinase transcripts.

To verify the *in vivo* relevance of the sequence variations detected in tumor cell lines, we analyzed cDNA from 165 primary breast, kidney, and prostate cancer specimens as well as

Gene	Mutation	Total number	Bladder	Bone and soft tissue	Brain	Breast	Cervix and vulva	Colon	Endometrium and placenta	Head and neck	Hematopoietic and lymphoid System
FGFR4	V10I	8		RD, I				LS-174T, I #	JAR, I #		OCI-AML5, I #
				TE-671, I				LS-180, I #				
FGFR4	L136P	37	SCaBER, P	SaOS2, P #		BT-474, P		HCT-116, P #			HL-60, P	
						BT-483, P #		LS-123, P #			K-562, P #	
						MB-157, P		LS-174T, P #			PLB-985, P	
						MB-415, P		LS-180, P #			U-266, P	
								LoVo, P				
								SW-837, P				
FGFR4	Y367C	1				MB-453, C						
FGFR4	G388R	58	TCCSUP, R		SHSY5Y, R #	Hs-578T, R	Ms 751, R	Caco2, R		SCC-4, R	EM-2, R	
					SK-N-SH, R #	MB-361, R #		SW-403, R #		SCC-22B, R	M-Mac-1, R #	
					SW-1088, R #	MB-453, R				SCC-22A, R	M-Mac-6, R #	
					U-1240, R	SK-BR-3, R						
					U-1242, R							
....												

Figure 2. Characterization of PTK genes with regard to genetic alterations detected in the transcripts of 276 tumor cell lines and control samples. As exemplified here for FGFR4, the spectrum of identified genetic alterations and the corresponding patterns of affected tumor cell lines or control samples was determined for each tyrosine kinase gene (Supplementary Table S8). The total sample number carrying a given sequence variant is indicated, and affected cancer cell lines are subdivided according to their tissue origin. *Blue*, germ line polymorphisms; *yellow*, somatic mutations. Heterozygosity is indicated by a hash; the other samples are homozygous carriers of the respective variation.

blood DNA from 90 healthy individuals for the occurrence of a representative subset of our identified genetic alterations. This subset was defined as all nonconservative sequence changes that were found at least twice in our panel of cell lines and control samples with at least one cell line originating from breast,

kidney, or prostate cancer. All but 2 of the 46 polymorphisms that fulfilled these criteria could be verified in patient sample cDNAs or blood DNA (Supplementary Table S11), hence confirming the *in vivo* relevance of sequence variations in tumor cell lines.

A						B															
Frequency	MS	NS	DEL	INS	Total	Protein domains and number of polymorphisms therein															
1x	36	0	2	1	39	Kinase	30	Cysteine rich	1	SH2	3										
2 - 5x	28	0	1	0	29	Transmembrane	3	Immunoglobulin	10	IPT/TIG	1										
6 - 10x	17	0	2	1	20	Juxtamembrane	13	Signal peptide	1	Frizzled	1										
> 10x	50	0	11	6	67	FN Type III	8	Sema	3	WIF	1										
Total	131	0	16	8	155	SAM	3	Proline rich	9	Other	67										

C		Tissue origin																			
Gene	Germ line Alteration	BL (5)	BS (4)	BA (16)	BE (22)	CV (11)	CO (23)	EP (3)	HN (13)	HL (24)	KI (9)	LI (4)	LU (22)	OV (11)	PA (17)	PR (6)	SK (53)	ST (5)	TE (3)	TY (3)	NO (22)
EGFR	R521K (SI 15)		1/2	3/16	3/22	2/11	12/23	1/3	9/13		7/9	2/4	5/22	4/11	5/17	2/6	8/52	2/6	2/2		12/22
FGFR4	G388R (SI 9)	1/4		5/14	4/11	1/8	2/13		3/12	3/13	3/8	2/3	6/17	2/5	4/9	2/6	11/45	1/1			8/19
MET	R988C (SI 10)												1/20								1/19
MET	T1010I (SI 11)			1/15	3/21		1/22			1/13						1/6	1/50				
NTRK1	R780Q (SI 5)						1/18		1/12					1/11							
ROR1	M518T	2/4	4/4	13/15	14/15	8/9	18/23	3/3	7/10	9/10	7/8	2/3	15/21	7/10	14/16	5/6	44/51	4/5	1/3		21/22
ROR2	T245A			1/1	5/5	2/4	3/3	1/2	1/1	1/3	2/7	4/4	1/1	2/4	3/3	1/1	3/4		1/1		3/8
TYRO3	I346N	2/4	3/4	9/15	1/9	3/8	4/16	3/3	3/12	7/23	2/8	1/4	7/22	3/6	7/12		18/46				12/21
JAK3	V722I (SI 7)								2/7												
TNKK1	M598delinsEVRSHX		2/2	12/15	1/22	1/10				13/21	4/9	1/4	4/20	1/10	1/17	1/6	21/38		1/3		1/21
TXK	R336Q	1/5			1/20		3/23		4/10			1/4		2/11			2/45	1/7			1/20
TXK	Y414fsX15			3/10	4/19	1/9	2/22			9/24	2/4		2/16		3/15		1/39	2/4		2/3	2/19
TYK2	E971fsX67				5/17	1/11	3/21	1/3					1/22		2/14				1/2		1/21
TYK2	V362F	1/4	4/4	9/12	2/4	2/9	5/12			14/21	6/9	1/4	7/20		2/9		7/24				4/13

Figure 3. Distributions of nonsynonymous polymorphisms identified in the TKT of 276 cancer cell lines and control samples. *A*, rates of germ line alterations. The rates of missense (MS) or nonsense (NS) substitutions, deletions (DEL) and insertions (INS) subdivided into four frequency categories (1, 2–5, 6–10, or >10 affected samples) are summarized. *B*, domain localization of identified polymorphisms. The number of polymorphisms detected in distinct domains or other protein regions is indicated. *C*, tissue distribution of germ line variations. The tissue distribution (BL, bladder; BS, bone and soft tissue; BA, brain; BE, breast; CV, cervix and vulva; CO, colon; EP, endometrium and placenta; HN, head and neck; HL, hematopoietic and lymphoid system; KI, kidney; LI, liver; LU, lung; OV, ovary; PA, pancreas; PR, prostate; SK, skin; ST, stomach; TE, testes; TY, thyroid; NO, normal control samples) was determined for all polymorphisms (Supplementary Table S10) and presented here for those described in the text. *Paired numerals*, number of carriers of the indicated variant as a subset of all cell lines with the same tissue origin that express the corresponding gene regardless of its genotype. *Bold type*, novel germ line alterations; *parenthesized numbers*, supplementary references that associate respective polymorphisms with cancer.

Cancer relevance of identified polymorphisms. Some of the identified polymorphisms have previously been associated with nonproliferative diseases. Respective functional modulations may, because of the pleiotropic effects of many PTKs, also be relevant for cancer. This is exemplified by the V722I transversion in the pseudokinase domain of Janus-activated kinase 3 that we identified as a rare heterozygous polymorphism in the head and neck cancer cell lines SCC-10A and SCC-10B (Fig. 3C). First reported in patients with autosomal recessive T-B+ SCID syndrome (19), its recent detection in an acute megakaryoblastic leukemia patient and the capacity to transform Ba/F3 cells (20) support a potential role in cancer. Another example is NTRK1 R780Q, which we found in the colon, ovarian, and head and neck cancer cell lines Caco2, SK-OV-8, and SCC-9 (Fig. 3C), respectively. This SNP affects the same arginine residue whose replacement with proline was shown to be associated with “congenital insensitivity to pain with anhidrosis” and abrogation of catalytic tyrosine kinase activity *in vitro* (21). Assuming a similar loss-of-function for the NTRK1 Q780 isotype, this variant may exert antiapoptotic and hence pro-oncogenic effects, as expression of NTRK1 wild-type was associated with induction of apoptosis and a favorable prognosis of neuroblastoma patients (22).

Cancer relevance was also established for MET T1010I, which represents a biomarker for MET inhibitor efficacy (23) and was originally reported as a somatic gain-of-function mutation in small cell lung cancer and non-small cell lung cancer (NSCLC; ref. 24) and malignant pleural mesotheliomas (23). Its detection in 4 of our 90 blood control DNAs (Supplementary Table S11), however, confirmed previous hints for potential germ line occurrence (25). Moreover, the identification of MET T1010I in the prostate carcinoma cell line TSU-PR1 and a primary prostate tumor as well as in the brain, breast, colon, hematopoietic, and skin cancer cell lines IHR-32, DAL, LS-123, U-266, and Colo-829, respectively (Fig. 3C and Supplementary Tables S10 and S11), suggests enhanced MET signaling in these tumor cell lines and expands the currently reported spectrum of affected tumor types.

Polymorphism frequencies in cancer cells versus normal tissues. Differential occurrence rates of sequence polymorphisms in particular cancer types and/or normal tissues may indicate tumor suppressive or promoting effects. To address the potential relevance of all polymorphism for certain tumor types, we compared their occurrence frequencies in tissue types and control samples (Fig. 3C and Supplementary Table S10). Only some examples are displayed in Fig. 4. For epidermal growth factor receptor (EGFR) R521K, a relative overrepresentation of the EGFR K521 allele in cDNAs of normal control samples (55%), colon (52%), and head and neck (69%) tumor cell lines was detected (Fig. 4, left). This indicates a possible tumor suppressive activity of the EGFR K521 isotype that apparently is not relevant to colon cancer and head and neck cancer. An attenuated growth response to EGFR ligands and reduced induction of the proto-oncogenes *FOS*, *JUN*, and *MYC* in EGFR K521, but not EGFR R521-expressing cells (26), and an increased risk of local recurrence after chemoradiation treatment for rectal cancer patients with at least one EGFR R521 allele (27) support these conclusions. Similarly, a clearly differential occurrence of TYK2 F362 allele carriers was observed in brain (75%)–derived and hematopoietic/lymphoid system (67%)–derived tumor cell lines compared with control tissues (31%) or other tumor types (Fig. 4, middle). The novel polymorphism TNK M598delinsEVRSHX was found at low frequencies in control samples (5%) and cancer cells of several tissue origins, but occurred in

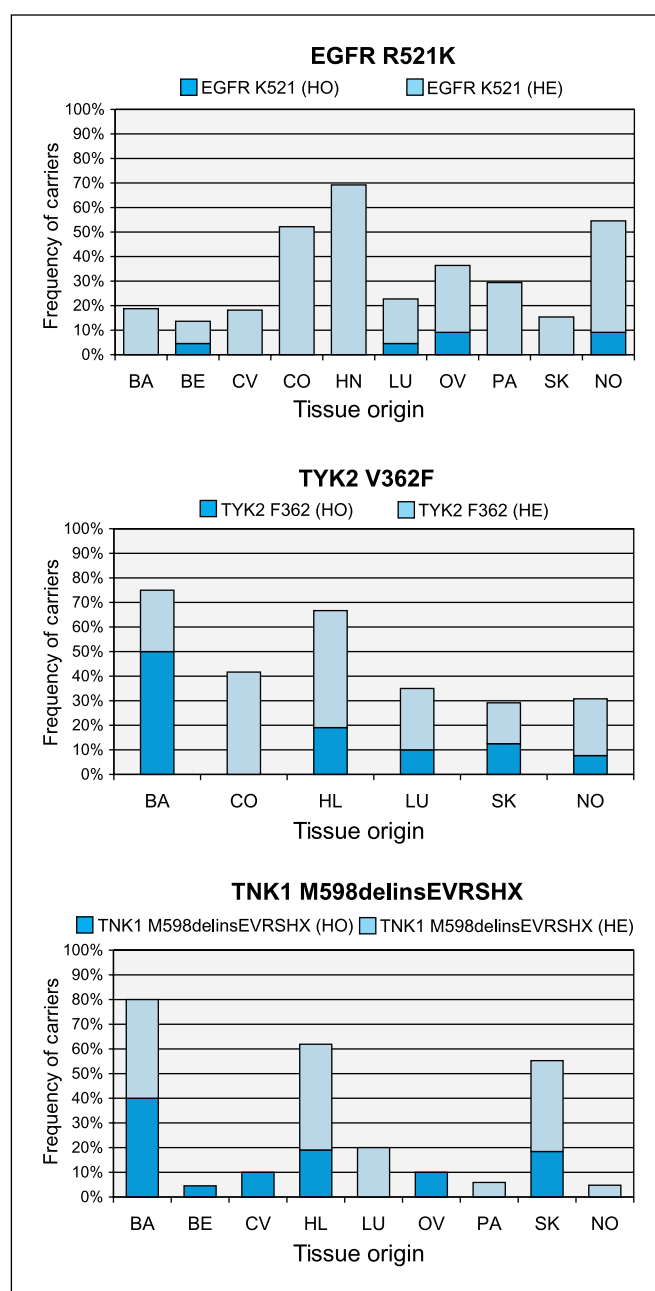


Figure 4. Diverging occurrence rates of polymorphisms in different tumor types and/or control samples. The frequency of homozygous (HO; dark blue column) and heterozygous (HE; light blue column) carriers of EGFR R521K, TYK2 V362F, and TNK1 M598delinsEVRSHX was determined. Only tissue origins (for abbreviations, see legend to Fig. 3) with an expression of the corresponding gene in at least 10 samples have been selected for this analysis.

62% of blood-derived, 55% of skin-derived, and, even more prominent, 80% of brain-derived tumor cell lines (Fig. 4, right). In contrast to EGFR R521K, the underrepresentation of the TYK2 F362 allele and the TNK insertion in control samples indicates a tumor-promoting function with particular relevance for leukemia, melanoma, and glioma. It can be expected that, as for EGFR R521K (27) or FGFR4 G388R (10), the correlation with clinical parameters will assign therapeutic and/or predictive value to many of such unequally distributed alleles.

Polymorphisms affecting the kinase domain. The localization of genetic alterations within the respective protein sequence may

be indicative of structural and/or functional consequences. The skipping of entire exons in the cytoplasmic kinases TYK2 and TXK, TYK2 E971fsX67 and TXK Y414fsX15, for instance, results in frame shifts and premature translation termination in the tyrosine kinase domains, and thus most likely is associated with catalytic inactivation. The truncated TYK2 variant that lacks 206 amino acids of the kinase domain, including the catalytic site and the activation loop, may also impose a dominant negative effect on cell signals. Interestingly, Stoiber et al. (28) reported that TYK2-deficient mice developed B and T lymphoid leukemia with higher incidence and shortened latency as a result of decreased cytotoxic capacity of TYK2^{-/-} natural killer (NK) and NKT cells and thus impaired tumor surveillance. Because NK activity as part of the innate immune system mediates tumor rejection in general, the significance of TYK2 loss-of-function might not be restricted to hematopoietic malignancies, but may also be important for other cancer types. Consistent with this possibility, we detected TYK2 E971fsX67 in cancer cells derived from various tissues, including breast, cervix/vulva, colon, endometrium, lung, and pancreas (Fig. 3C), as well as 33 clinical breast, prostate, and kidney cancer specimens. Its occurrence in the control cell line BPH-1 suggests potential germ line origin. Hence, the TYK2 E971fsX67 splice variant may also represent a prognostic marker for cancer patients and support therapeutic decision making.

Overall, these examples point at the potential role of sequence polymorphisms as genetic parameters that may contribute to a patient-specific definition of disease predisposition, rate of progression, or responsiveness to therapeutic agents. In conjunction with simple detectability in blood samples, this renders

polymorphisms to be highly valuable biomarkers for diagnostic patient characterization.

Identification of 234 somatic PTK gene mutations. Of all sequence differences, 234 were undetectable in any of the control samples or public databases and were thus defined as somatic mutations. However, because of the lack of cell line-specific normal tissue controls, we cannot exclude the possibility that some actually represent rare germ line polymorphisms. The somatic mutations are composed of 210 missense and 2 nonsense single nucleotide substitutions as well as 19 deletions and 3 insertions. Although the majority (186) occurred once, 53 were found two to five times, and 3 in 6 to 10 tumor cell lines (Fig. 5A). Among the twice-occurring somatic mutations, 20 were detected in cell lines originating from the same tumor donor (Supplementary Table S1). They may be considered single mutations, thus adding up to a total of 206 nonrecurring mutational events. As for the polymorphisms, we present all somatic TKT alterations in the context of the respective protein domains and tumor types, and display the ratio of affected and expression-positive cell lines for each tissue origin (Fig. 5B and C; Supplementary Table S12).

Consistent with SYK A353T to represent one of the two most prevalent mutations, the tumor-suppressive tyrosine kinase SYK turned out to be the most frequently mutated kinase within our panel of 254 tumor cell lines. When absolute numbers of somatic mutations were compared, SYK scored highest with mutations detected in 11 nonrelated tumor cell lines (Supplementary Table S13). After normalization with respect to the PTK transcription status, SYK showed the highest mutation rate of 30.3 sporadic alterations per 1 Mb expressed coding sequence, followed by

A

Frequency	MS	NS	DEL	INS	Total
1x	171	0	7	2	180
2 - 5x	38	2	11	1	52
6 - 10x	1	0	1	0	2
> 10x	0	0	0	0	0
Total	210	2	19	3	234

B

Protein domains and number of somatic mutations therein					
Kinase	71	Cysteine-rich	6	Src-homology 2	4
Transmembrane	4	Immunoglobulin	7	Src-homology 3	3
Juxtamembrane	17	Signal peptide	2	Other	95
FN type III	10	Sema	1		
SAM	5	Proline-rich	9		

C

Gene	Somatic Mutations	Tissue origin																	
		BL (5)	BS (4)	BA (16)	BE (22)	CV (11)	CO (23)	EP (3)	HN (13)	HL (24)	KI (9)	LI (4)	LU (22)	OV (11)	PA (17)	PR (6)	SK (53)	ST (5)	TE (3)
EGFR	G719S (SI 17)						1/21												
EPHB2	Q722X (SI 18)															2/6			
FGFR1	P252S																1/38		
FGFR4	Y367C				1/14														
FLT3	R849H														1/2				
KIT	N822K (SI 19)								1/13										
MET	D981_E1027del (SI 20)				1/22								1/22				1/5		
RON	A1022_K1090del				1/17														
TEK	A1006T																1/31		
TYRO3	E489K				1/8										2/5				
ABL1	G417E																1/51		
ARG	K450R									1/9									
CSK	Q26X						2/18												
TEC	W531R									1/24									
YES1	K113Q				1/9		1/19				1/9								

Figure 5. Distributions of nonsynonymous somatic mutations identified in all transcribed PTK genes from 254 tumor cell lines. **A**, rates of somatic mutations. The allocation to missense or nonsense substitutions, deletions and insertions as well as frequency categories (1, 2–5, 6–10, or >10 affected samples) is shown. **B**, domain localization of identified mutations. The localization within defined domains or other protein regions is indicated. **C**, tissue distribution of sporadic alterations. For each somatic mutation, the tissue distribution (see legend to Fig. 3 for abbreviations) was determined (Supplementary Table S12) and presented here for text-related examples. *Paired numerals*, the number of mutated and expression-positive cell lines within a given tumor type. *Bold type*, novel somatic mutations; *numbers in parenthesis*, supplementary references that associate respective mutations with cancer.

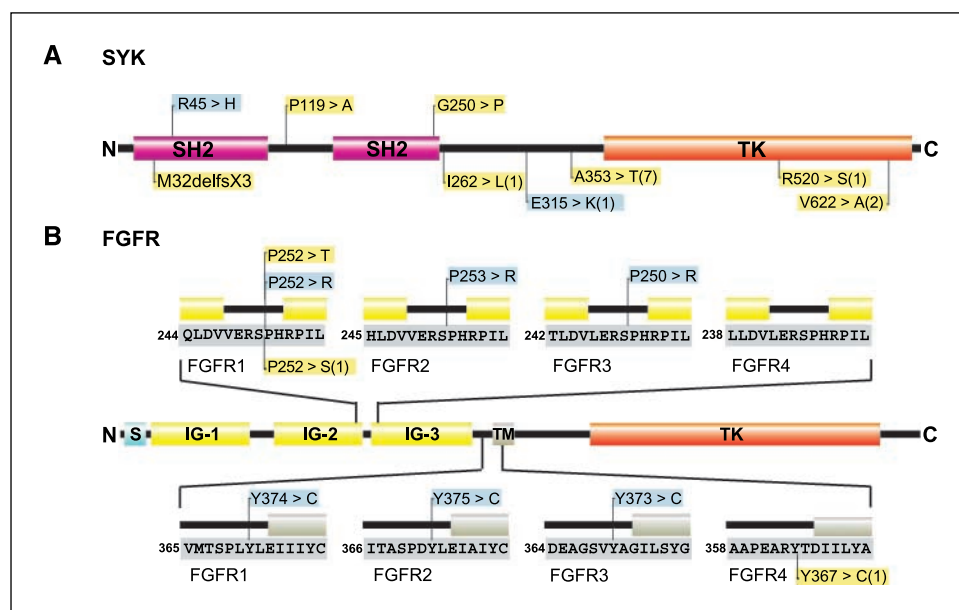


Figure 6. Illustration of known and novel genetic alterations in selected genes.

A, SYK. The domain organization and location of genetic alterations is displayed. **B, sequence comparison of FGFR1-4.** For FGFR1-4, the general domain organization (*middle*) and sequence comparisons of the linker region connecting the IG-D2-domain and IG-D3-domain (*top*) as well as a part of the extracellular juxtamembrane region (*bottom*) are illustrated. Genetic alterations identified in our cell line screen are indicated below, whereas known sequence variants are depicted above the graphical representation of the domain structure. *Blue*, polymorphisms; *yellow*, somatic mutations. *Numbers in parenthesis*, number of affected nonrelated cell lines (SH2, Src homology 2 domain; TK, tyrosine kinase domain; S, signal peptide; TM, transmembrane domain; IG, immunoglobulin-like domain).

NTRK1, *EPHA2*, and *FLT3* (Supplementary Table S14). The domain organization of SYK and the known and novel genetic alterations are shown in Fig. 6A.

Somatic mutations with possible oncogenic potential. Somatic mutations clustering in the EGFR kinase domain (EGFR G719S, L858R, L861Q, and others) have recently been reported for patients with gefitinib-responsive NSCLC and were shown to enhance tyrosine kinase activity and sensitivity to gefitinib *in vitro* (29–31). We found the EGFR G719S mutation to be heterozygously expressed in the colon cancer cell line SW-48 (Fig. 5C; Supplementary Tables S1 and S2). This shows the existence of Iressa sensitivity-mediating mutations in cancers other than NSCLC and, in particular, suggests colon cancer as another potential indication for gefitinib therapy.

Similar to EGFR L858R and gefitinib, the KIT N822K mutation that we confirmed in the acute myelogenous leukemia cell line KASUMI-1 (32) was reported to mediate sensitivity to Gleevec (33). The enhanced *in vitro* receptor activation shown for these EGFR and KIT mutations (30, 34) might be related to their location within the regulatory activation loop. In this respect, the sporadic variations FLT3 R849H, TEK A1006T, ABL G417E, ARG K450R, and TEC W531R, which we detected homozygously or heterozygously in BM-1604, SK-MEL-2, MM-Leh, Caki-2, and Jurkat (Fig. 5C), are particularly intriguing as they are located in the activation loop as well. By inference, these mutations may also have a higher probability to modulate the tyrosine kinase catalytic activity and/or related signaling pathways within the respective tumor cell lines.

The 18 somatic mutations we identified in intracellular juxtamembrane domains (Supplementary Table S12) might affect functionally important elements that mediate down-regulation of RTK activity. The in-frame deletion MET D981_E1027del as a result of exon 14 skipping, for instance, leads to the loss of c-Cbl E3-ligase binding, decreased ubiquitination, and prolonged ligand-dependent cell signaling *in vitro* and *in vivo* (35). Although we confirmed MET D981_E1027del in the NSCLC cell line NCI-H596, its homozygous detection in breast and stomach cancer cell lines

MDA-MB-415 and Hs746T, respectively (Fig. 5C), provides evidence for its occurrence in tumor types other than the reported NSCLC (35, 36). Presuming enhanced sensitivity to anti-MET therapeutics that MET D981_E1027del was suggested to mediate (35), our findings extend the potential clinical relevance for this deletion.

The somatic mutations we identified within the extracellular domain of two FGFR family members, FGFR1 P252S and FGFR4 Y367C (Figs. 5C and 6B), possibly augment receptor activation by influencing ligand binding and receptor dimerization, respectively. The highly conserved FGFR1 P252 residue we found to be heterozygously exchanged with hydrophilic serine in the melanoma cell line MeWo has previously been shown to be replaced by threonine in lung cancer (2) and arginine in patients with Pfeiffer syndrome (37). The crystal structure of the homologous activating FGFR2 mutant, FGFR2 P252R, revealed the formation of three additional hydrogen bonds with complexed fibroblast growth factor 2 (FGF2). They were predicted to increase the receptor affinity for its specific ligand as well as to allow binding of a different set of ligands (38). Because the hydroxyl group of the P252-replacing serine residue in FGFR1 also has a high potential to form additional hydrogen bonds, the somatic FGFR1 P252S substitution may represent a gain-of-function mutation with analogous functional consequences as for FGFR2 P252R. This is particularly intriguing in the context of studies demonstrating that blockage of FGFR1 or basic fibroblast growth factor function was associated with suppressed proliferation and survival of melanoma cells (39).

The novel FGFR4 Y367C mutation was detected as a homozygous genotype in the breast cancer cell line MDA-MB-453, and the affected Y367 residue in the extracellular juxtamembrane domain is highly conserved throughout the FGF family. Remarkably, homologous substitutions in FGFR1 (Y372C), FGFR2 (Y375C), and FGFR3 (Y373C) were shown to cause various osteogenic deficiency syndromes (40) through the formation of intermolecular disulfide bonds that force receptor dimerization and activation. Ligand-independent, constitutive receptor activation has been confirmed *in vitro* for FGFR1 Y372C (41) and FGFR3 Y373C (42). Furthermore,

the oncogenic potential of the FGFR3 Y373C variant has been shown and was suggested to contribute to tumor progression of multiple myeloma (43). Thus, it is most likely that Y367C as the homologous FGFR4 variant also results in basal receptor activation, which strongly indicates an important role of this mutant in cancer.

Nonsense substitutions abrogate tumor suppressor activity. Down-regulation of tumor-suppressive activity is expected for the two nonsense substitutions we detected in *EPHB2* and *CSK* (Fig. 5C). Q722X-mediated truncation and kinase inactivation of *EPHB2* in the two prostate cancer cell lines BM-1604 and DU-145 supports mutational inactivation to be involved in progression of prostate cancer as proposed by Huusko et al. (44). They showed suppressed growth and colony formation of DU-145 cells upon reconstitution with functional *EPHB2*. The heterozygous *CSK* Q26X nonsense substitution detected in the colon cancer cell lines DLD-1 and HCT-15 is consistent with reduced protein levels of this negative regulator of SRC family kinases that were reported for ~60% of human colon cancer cases with elevated SRC activity (45). These data indicate a significant role of *CSK* nonsense mutations in the development and/or progression of colon carcinoma and therefore strongly suggest the inclusion of SRC kinase inhibitors in the therapeutic regimen of this prevalent malignancy.

The examples discussed above represent only a partial extract of our overall data. Other genetic alterations affecting less investigated PTKs such as members of the AATYK, DDR, EPH, ROR, ROS, or FRK families, as well as tyrosine kinases that more recently captured scientific attention such as HER3 or ACK1, have been found (Supplementary Tables S2–S6). Their identification will support novel functional investigations toward the understanding of the therapeutic value of these kinases.

Low redundancy of PTK gene mutations in human tumors. In agreement with results from previous studies (1–7, 46), our analysis of 254 cancer cell lines and additional primary tumors indicates that mutational patterns might be quite unique for the majority of human tumors, and that the frequency of specific somatic mutations in PTKs is low. Data mining of public databases and the literature revealed that only nine of all sporadic alterations identified in our study have been described before (Supplementary Table S12). Among them are KIT N822K and VEGFR1 R781Q as the only two alterations that were picked up in the currently most

comprehensive mutational kinome analysis of human cancer samples (7). The low redundancy of somatic mutations is furthermore reflected by the nonrecurrence that we found for 206 of the 234 mutational events within our panel of tumor cell lines.

Consistent with this picture, none of the seven somatic representatives of our exemplary subset of nonconservative and more frequent alterations found in at least one breast, kidney, or prostate cancer cell line could be detected in any of the 165 primary breast, kidney, and prostate cancer specimens (Supplementary Table S11). In fact, two genetic alterations, YES K113Q and TYRO3 E489K, were found in blood controls and therefore must be considered rare germ line alterations.

Despite the low redundancy of individual mutations, 70 of the tyrosine kinase genes turned out to carry at least one somatic mutation. Although most of our mutations require further experimental evaluation to determine their cancer relevance and in some cases may turn out to represent “passenger” rather than “driver” mutations, this broad incidence of sporadic alterations underscores the central importance of the entire PTK family in oncogenesis. Moreover, it provides further compelling support for the development of multitargeted kinase inhibitors or combination of complementary therapeutics as cancer treatments that may be adapted to the pathologic and genetic parameters of an individual patient. The extensive characterization of established tumor cell lines with respect to transcriptional profiles of genetic variations in this currently most promising cancer target family will aid in the selection of suitable cell systems, data interpretation and target validation, and thereby support preclinical development of novel targeted cancer drugs.

Acknowledgments

Received 7/17/2007; revised 9/21/2007; accepted 10/9/2007.

Grant support: Biomedical Research Council of A*STAR, Singapore, and the Max-Planck Society, Germany.

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.

We thank Gunaretnam Rajagopal and Lim Yun Ping from A*STAR BII for strong support; the BioComputing Center at BII and Chew Ah Keng from the DNA sequencing facility at the Singapore Institute of Molecular and Cell Biology for expert assistance; and Philip Yeo and Hwai Loong Kong, without whom this project would not have been possible.

References

- Stephens P, Edkins S, Davies H, et al. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet* 2005;37:590–2.
- Davies H, Hunter C, Smith R, et al. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res* 2005;65:7591–5.
- Bignell G, Smith R, Hunter C, et al. Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults. *Genes Chromosomes Cancer* 2006;45:42–6.
- Stephens P, Hunter C, Bignell G, et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* 2004;431:525–6.
- Bardelli A, Parsons DW, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 2003;300:949.
- Thomas RK, Baker AC, Debiassi RM, et al. High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 2007;39:347–51.
- Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007;446:153–8.
- Balakrishnan A, Bleeker FE, Lamba S, et al. Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. *Cancer Res* 2007;67:3545–50.
- Hunter K. Host genetics influence tumour metastasis. *Nat Rev Cancer* 2006;6:141–6.
- Bange J, Prechtel D, Cheburkin Y, et al. Cancer progression and tumor cell motility are associated with the FGFR4 Arg(388) allele. *Cancer Res* 2002;62:840–7.
- Thussbas C, Nahrig J, Streit S, et al. FGFR4 Arg388 allele is associated with resistance to adjuvant therapy in primary breast cancer. *J Clin Oncol* 2006;24:3747–55.
- Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 2007;39:352–8.
- Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–93.
- Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet* 2007;3:e42.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* 2002;298:1912–34.
- Puissant C, Houdebine LM. An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Bio-techniques* 1990;8:148–9.
- Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. 2nd ed. New York: Cold Spring Harbor Press; 1989.
- Ikedobi ON, Davies H, Bignell G, et al. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther* 2006;5:2606–12.

19. Schumacher RF, Mella P, Badolato R, et al. Complete genomic organization of the human JAK3 gene and mutation analysis in severe combined immunodeficiency by single-strand conformation polymorphism. *Hum Genet* 2000;106:73–9.
20. Walters DK, Mercher T, Gu TL, et al. Activating alleles of JAK3 in acute megakaryoblastic leukemia. *Cancer Cell* 2006;10:65–75.
21. Greco A, Villa R, Tubino B, Romano L, Penso D, Pierotti MA. A novel NTRK1 mutation associated with congenital insensitivity to pain with anhidrosis. *Am J Hum Genet* 1999;64:1207–10.
22. Lavoie JF, Lesauteur L, Kohn J, et al. TrkA induces apoptosis of neuroblastoma cells and does so via a p53-dependent mechanism. *J Biol Chem* 2005;280:29199–207.
23. Jagadeeswaran R, Ma PC, Seiwert TY, et al. Functional analysis of c-Met/hepatocyte growth factor pathway in malignant pleural mesothelioma. *Cancer Res* 2006;66:352–61.
24. Ma PC, Kijima T, Maulik G, et al. c-MET mutational analysis in small cell lung cancer: novel juxtamembrane domain mutations regulating cytoskeletal functions. *Cancer Res* 2003;63:6272–81.
25. Tengs T, Lee JC, Paez JG, et al. A transforming MET mutation discovered in non-small cell lung cancer using microarray-based resequencing. *Cancer Lett* 2006;239:227–33.
26. Moriai T, Kobrin MS, Hope C, Speck L, Korc M. A variant epidermal growth factor receptor exhibits altered type α transforming growth factor binding and transmembrane signaling. *Proc Natl Acad Sci U S A* 1994;91:10217–21.
27. Zhang W, Park DJ, Lu B, et al. Epidermal growth factor receptor gene polymorphisms predict pelvic recurrence in patients with rectal cancer treated with chemoradiation. *Clin Cancer Res* 2005;11:600–5.
28. Stoiber D, Kovacic B, Schuster C, et al. TYK2 is a key regulator of the surveillance of B lymphoid tumors. *J Clin Invest* 2004;114:1650–8.
29. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–500.
30. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–39.
31. Pao W, Miller V, Zakowski M, et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A* 2004;101:13306–11.
32. Beghini A, Magnani I, Ripamonti CB, Larizza L. Amplification of a novel c-Kit activating mutation Asn(822)-Lys in the Kasumi-1 cell line: a t(8;21)-Kit mutant model for acute myeloid leukemia. *Hematol J* 2002;3:157–63.
33. Heinrich MC, Corless CL, Demetri GD, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol* 2003;21:4342–9.
34. Kemmer K, Corless CL, Fletcher JA, et al. KIT mutations are common in testicular seminomas. *Am J Pathol* 2004;164:305–13.
35. Kong-Beltran M, Seshagiri S, Zha J, et al. Somatic mutations lead to an oncogenic deletion of met in lung cancer. *Cancer Res* 2006;66:283–9.
36. Ma PC, Jagadeeswaran R, Jagadeesh S, et al. Functional expression and mutations of c-Met and its therapeutic inhibition with SU11274 and small interfering RNA in non-small cell lung cancer. *Cancer Res* 2005;65:1479–88.
37. Muenke M, Schell U, Hehr A, et al. A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat Genet* 1994;8:269–74.
38. Ibrahimi OA, Eliseenkova AV, Plotnikov AN, Yu K, Ornitz DM, Mohammadi M. Structural basis for fibroblast growth factor receptor 2 activation in Apert syndrome. *Proc Natl Acad Sci U S A* 2001;98:7182–7.
39. Wang Y, Becker D. Antisense targeting of basic fibroblast growth factor and fibroblast growth factor receptor-1 in human melanomas blocks intratumoral angiogenesis and tumor growth. *Nat Med* 1997;3:887–93.
40. Wilkie AO. Bad bones, absent smell, selfish testes: the pleiotropic consequences of human FGF receptor mutations. *Cytokine Growth Factor Rev* 2005;16:187–203.
41. White KE, Cabral JM, Davis SI, et al. Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation. *Am J Hum Genet* 2005;76:361–7.
42. d’Avis PY, Robertson SC, Meyer AN, Bardwell WM, Webster MK, Donoghue DJ. Constitutive activation of fibroblast growth factor receptor 3 by mutations responsible for the lethal skeletal dysplasia thanatophoric dysplasia type I. *Cell Growth Differ* 1998;9:71–8.
43. Chesi M, Brents LA, Ely SA, et al. Activated fibroblast growth factor receptor 3 is an oncogene that contributes to tumor progression in multiple myeloma. *Blood* 2001;97:729–36.
44. Huusko P, Ponciano-Jackson D, Wolf M, et al. Nonsense-mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. *Nat Genet* 2004;36:979–83.
45. Rengifo-Cam W, Konishi A, Morishita N, et al. Csk defines the ability of integrin-mediated cell adhesion and migration in human colon cancer cells: implication for a potential role in cancer metastasis. *Oncogene* 2004;23:289–97.
46. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314:268–74.