

# Sixteen–Kinase Gene Expression Identifies Luminal Breast Cancers with Poor Prognosis

Pascal Finetti,<sup>1</sup> Nathalie Cervera,<sup>1</sup> Emmanuelle Charafe-Jauffret,<sup>1,2,5</sup> Christian Chabannon,<sup>3</sup> Colette Charpin,<sup>5,6</sup> Max Chaffanet,<sup>1</sup> Jocelyne Jacquemier,<sup>1,2</sup> Patrice Viens,<sup>4</sup> Daniel Birnbaum,<sup>1</sup> and François Bertucci<sup>1,4,5</sup>

<sup>1</sup>UMR599 Inserm, Institut Paoli-Calmettes, Laboratoire d'Oncologie Moléculaire, Centre de Recherche en Cancérologie de Marseille,

<sup>2</sup>Département de BioPathologie, <sup>3</sup>Centre de Ressources Biologiques, <sup>4</sup>Département d'Oncologie Médicale, Institut Paoli-Calmettes,

<sup>5</sup>Faculté de Médecine, Université de la Méditerranée, and <sup>6</sup>Département d'Anatomopathologie, Hôpital Nord, Marseille, France

## Abstract

**Breast cancer is a heterogeneous disease made of various molecular subtypes with different prognosis. However, evolution remains difficult to predict within some subtypes, such as luminal A, and treatment is not as adapted as it should be. Refinement of prognostic classification and identification of new therapeutic targets are needed. Using oligonucleotide microarrays, we profiled 227 breast cancers. We focused our analysis on two major breast cancer subtypes with opposite prognosis, luminal A ( $n = 80$ ) and basal ( $n = 58$ ), and on genes encoding protein kinases. Whole-kinome expression separated luminal A and basal tumors. The expression (measured by a kinase score) of 16 genes encoding serine/threonine kinases involved in mitosis distinguished two subgroups of luminal A tumors: Aa, of good prognosis and Ab, of poor prognosis. This classification and its prognostic effect were validated in 276 luminal A cases from three independent series profiled across different microarray platforms. The classification outperformed the current prognostic factors in univariate and multivariate analyses in both training and validation sets. The luminal Ab subgroup, characterized by high mitotic activity compared with luminal Aa tumors, displayed clinical characteristics and a kinase score intermediate between the luminal Aa subgroup and the luminal B subtype, suggesting a continuum in luminal tumors. Some of the mitotic kinases of the signature represent therapeutic targets under investigation. The identification of luminal A cases of poor prognosis should help select appropriate treatment, whereas the identification of a relevant kinase set provides potential targets.** [Cancer Res 2008;68(3):767–76]

## Introduction

Breast cancer is a heterogeneous disease whose clinical outcome is difficult to predict and treatment is not as adapted as it should be. Breast cancer can be defined at the clinical, histologic, cellular, and molecular levels. Efforts to integrate all these definitions improve our understanding of the disease and its management (1). Initial studies using DNA microarrays have identified five major breast cancer molecular subtypes (luminal A and B, basal, ERBB2 overexpressing, and normal-like; refs. 2–5).

These subtypes, which are defined by the specific expression of an intrinsic set of ~500 genes, are variably associated with different histologic types and with different prognosis. Luminal A breast cancers, which express hormone receptors, have an overall good prognosis and can be treated by hormone therapy. ERBB2-overexpressing breast cancers, which overexpress the ERBB2 tyrosine kinase receptor, have a poor prognosis and can be treated by targeted therapy using trastuzumab or lapatinib (6, 7). No specific therapy is available against the other subtypes, although the prognosis of basal and luminal B tumors is poor. This biologically relevant taxonomy remains imperfect because clinical outcome may be variable within each subtype, suggesting the existence of unrecognized subgroups.

Progress can be made in several directions. First, it is necessary to identify among good prognosis tumors, such as luminal A breast cancers, the ones that will relapse and metastasize. Second, a better definition of poor prognosis breast cancers and associated target genes will allow the development of new drugs that will in turn allow a better management of these cancers. We have here established the gene expression profiles of a series of breast tumor samples. To provide clues on both potential new prognostic and therapeutic targets, we have specifically focused our analysis on two major breast cancer subtypes with opposite prognosis (luminal A and basal) and on genes encoding protein kinases.

The human kinome constitutes about 1.7% of all human genes (8) and represents a great part of genes whose alteration contributes to oncogenesis (9). Protein kinases mediate most signal transduction pathways in human cells and play a role in most key cell processes. Some kinases are activated or overexpressed in cancers and constitute targets for successful therapies (10). In parallel to ongoing systematic sequencing projects (11), analysis of differential expression of kinases in cancers may identify new oncogenic activation pathways. As such, kinases represent an attractive focus for expression profiling in two important subtypes of breast cancer.

## Materials and Methods

**Patients and samples.** A total of 227 pretreatment early breast cancer samples were available for RNA profiling on Affymetrix microarrays. They were collected from 226 patients with invasive adenocarcinoma who underwent initial surgery at the Institut Paoli-Calmettes and Hôpital Nord (Marseille) between 1992 and 2004. Samples were macrodissected by pathologists and frozen within 30 min of removal in liquid nitrogen. All profiled specimens contained >60% of tumor cells. Characteristics of samples and treatment are listed in Supplementary Table S1.

In addition, we profiled RNA extracted from eight cell lines that provided models for cell types encountered in mammary tissues: three luminal

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

**Requests for reprints:** Daniel Birnbaum, UMR599 Inserm, 27 Bd. Leï Roure, 13009 Marseille, France. Phone: 33-4-91-75-84-07; Fax: 33-4-91-26-03-64; E-mail: birnbaum@marseille.inserm.fr.

©2008 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-07-5516

epithelial cell lines (HCC1500, MDA-MB-134, and ZR-75-30), three basal epithelial cell lines (HME-1, HMEC-derived 184B5, and MDA-MB-231), and two lymphocytic B- and T-cell lines (Daudi and Jurkatt, respectively). All cell lines were obtained from American Type Culture Collection<sup>7</sup> and were grown as recommended.

**Gene expression profiling with DNA microarrays.** Gene expression analyses were done with Affymetrix U133 Plus 2.0 human oligonucleotide microarrays containing >47,000 transcripts and variants, including 38,500 well-characterized human genes. Preparation of cRNA from 3 µg total RNA, hybridizations, washes, and detection were done as recommended by the supplier.<sup>8</sup> Scanning was done with Affymetrix GeneArray scanner, and quantification was done with Affymetrix GCOS software. Hybridization images were inspected for artifacts.

**Gene expression data analysis.** Expression data were analyzed by the Robust Multichip Average method in R using Bioconductor and associated packages (12). Before analysis, a filtering process removed the genes with low and poorly measured expression from the dataset as defined by expression value inferior to 100 units in all 227 breast cancer tissue samples, retaining 31,189 genes/ESTs.

Before unsupervised hierarchical clustering, a second filter excluded genes showing low expression variation across the 227 samples, as defined by SD inferior to 0.5 log<sub>2</sub> units (only for calculation of SD, values were floored to 100 because discrimination of expression variation in this low range cannot be done with confidence), retaining 14,486 genes/ESTs. Data were then log<sub>2</sub> transformed and submitted to the Cluster program (13) using data median centered on genes, Pearson correlation as similarity metric, and centroid linkage clustering. Results were displayed using TreeView program (13). Quality threshold clustering identifies sets of genes with highly correlated expression patterns among the hierarchical clustering. It was applied to the kinase probe sets and basal and luminal A tumors using TreeView program (13). The cutoffs for minimal cluster size and minimal correlation were 15 and 0.7, respectively. The gene clusters were interrogated using Ingenuity software to assess significant representation of biological pathways and functions.

**Definition of kinase-encoding probe sets.** The kinome database established by Manning et al. (8) was used as reference to extract the kinase-encoding genes from the Affymetrix Genechip U133 Plus 2.0. First, because annotation of the Human Genome Organization symbols did not correspond necessarily between the genes represented on the Affymetrix chip and the kinome, we used the mRNA accession number as cross-reference. cDNA sequences of the kinome were compared with the representative mRNA sequences of the Unigene database using BLASTn, and alignments between these sequences were obtained. All mRNAs with exact match were retained, and their accession numbers were compared with those of the 31,189 selected probe sets given by Affymetrix. Second, some kinase genes were represented by several probe sets on the Affymetrix chip. This may introduce bias in the weight of the groups of genes for analysis by quality threshold clustering. In these cases, probe sets with an extension «\_at», next «s\_at», and followed by all other extensions were preferentially kept. When several probe sets with the best extension were available, the one with the highest median value was retained. From the initial list of 518 kinases, we finally retained 435 probe sets representing 435 kinase genes (Supplementary Table S2).

**Collection of published datasets.** To test the performance of our multigene signature in other breast cancer samples, we analyzed three major publicly available data sets: van de Vijver et al. (14),<sup>9</sup> Wang et al. (15) collected from National Center for Biotechnology Information (NCBI)/Genbank GEO database (series entry GSE2034), and Loi et al. (16) collected from NCBI/Genbank GEO database (series entry GSE6532). Analysis of each data set was done in several successive steps: identification of molecular subtypes based on the common intrinsic gene set, identification of the kinase gene set common with ours, followed by computing of the kinase

score (see below) for the luminal A samples. Clinical data of luminal A samples from our series and public series used for analyses are detailed in Supplementary Table S3.

**Statistical analyses.** We defined a score, called the kinase score, which was based on the expression level of 16 kinase genes. It was defined as

$$KS = \frac{A}{n} \sum_{i=1}^n (x_i - B)$$

where *KS* is the kinase score; *A* and *B* represent normalization variables, which make the kinase score comparable across the different datasets; *n* as the number of available kinase genes (7–16); and *x<sub>i</sub>* as the logarithmic gene expression level in tumor *i*. Using a cutoff value of 0, each tumor was assigned a low score (kinase score < 0; i.e., with overall low expression of 16 kinase genes) or a high score (kinase score > 0; i.e., with overall strong expression of 16 kinase genes).

The samples included in the statistical analysis (luminal A subtype) were estrogen receptor (ER) and/or progesterone receptor (PR) positive as defined by using immunohistochemistry. We introduced two qualitative variables based on the mRNA expression level of ER and PR (*ESR1* probe set 205225\_at and *PGR* probe set 208305\_at): the cutoff for defining *ESR1* or *PGR* rich or *PGR* poor was the median expression level of the corresponding probe set. The two probe sets were chosen by using the same above-cited criteria.

Correlations between sample groups and histoclinical factors were calculated with the Fisher's exact test for qualitative variables with discrete categories and the Wilcoxon test for continuous variables. Follow-up was measured from the date of diagnosis to the date of last news for patients without relapse. Relapse-free survival (RFS) was calculated from the date of diagnosis until date of first relapse, whatever its location (local, regional, or distant) using the Kaplan-Meier method, and compared between groups with the log-rank test. The univariate and multivariate analyses were done using Cox regression analysis. The *P* values were based on log-rank test, and patients with one or more missing data were excluded. All statistical tests were two sided at the 5% level of significance. Statistical analysis was done using the survival package (version 2.30) in the R software (version 2.4.1).<sup>10</sup>

## Results

**Gene expression profiling of breast cancer and molecular subtypes.** A total of 227 samples were profiled using whole-genome DNA microarrays. Hierarchical clustering was applied to the 14,486 genes/ESTs with significant variation in expression level across all samples (Supplementary Fig. S1). Clusters of samples and clusters of genes were identified and represented previously recognized groups (17). We looked whether the five molecular subtypes reported by others (2–4) were also present in our series of samples by using the 476 genes common to the intrinsic 500-gene set. We had previously shown that clustering of the available RNA expression data for these 476 genes in the 122 samples from Sorlie et al. (17) discriminated the same five molecular subtypes, allowing the definition of typical expression profile of each subtype for our gene set (thereafter designated centroid) with 96% of concordance with those defined on the whole intrinsic gene set. We measured the Pearson correlation of each of our 227 tissue samples with each centroid. The highest coefficient defined the subtype, with a minimum threshold of 0.15. Subtypes are color coded in Supplementary Fig. S1: they included 91 luminal A samples and 67 basal samples as well as other subtypes.

**Whole-kinome expression profiling separates basal and luminal A breast cancers.** We wanted to identify kinase genes whose differential expression is associated with clinical outcome.

<sup>7</sup> <http://www.atcc.org/>

<sup>8</sup> <http://www.Affymetrix.com>

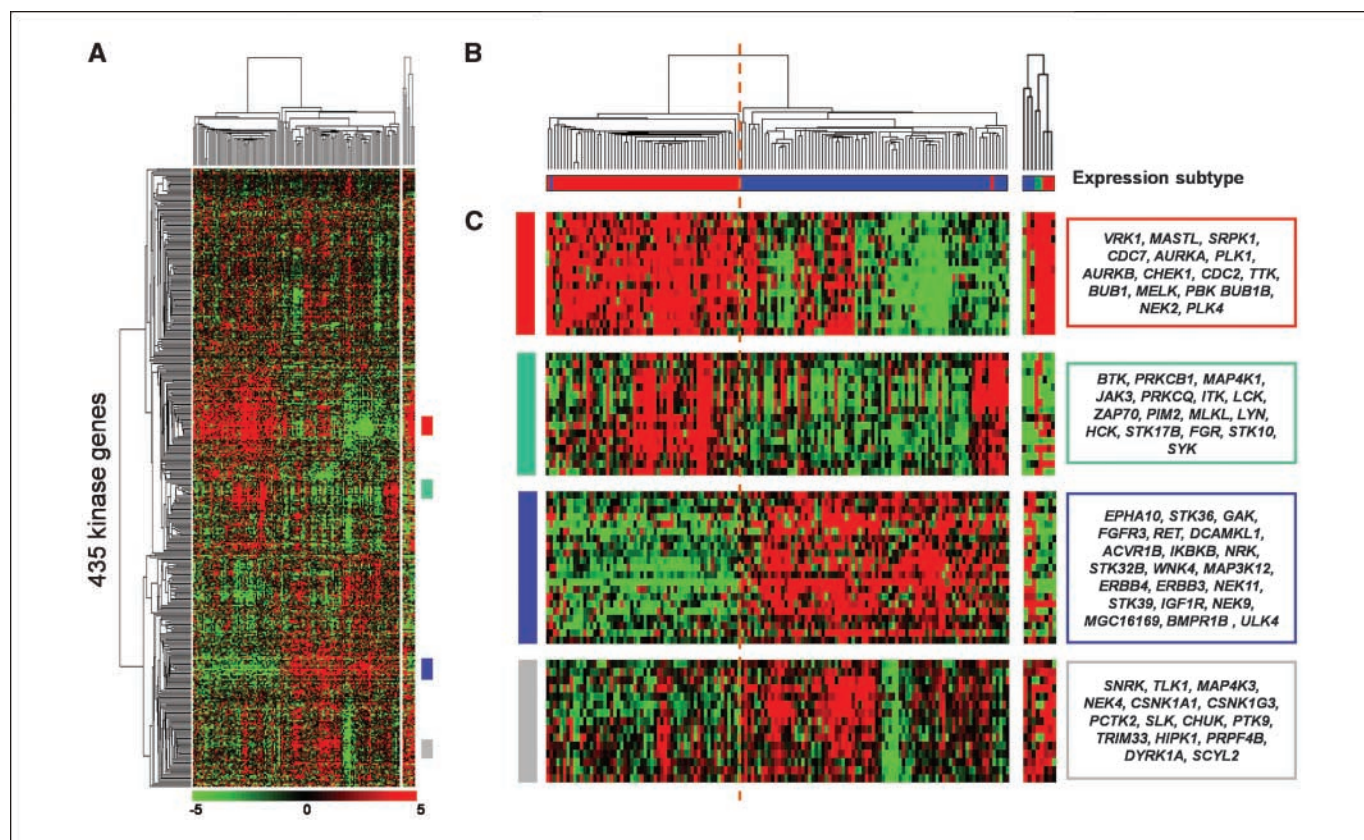
<sup>9</sup> Collected from [http://microarrays-pub.stanford.edu/wound\\_NKI/](http://microarrays-pub.stanford.edu/wound_NKI/).

<sup>10</sup> <http://www.cran.r-project.org>

We focused our analysis on two major subtypes of breast cancer with opposite prognosis, the basal and the luminal A subtypes. From our subtyping, we selected a series of 138 breast cancer samples with available full histoclinical annotations, including 80 luminal A and 58 basal breast cancers. We identified a total of 435 unique Affymetrix probe sets for 435 kinases as satisfying simultaneously presence, quality, and reliability (Supplementary Tables S2 and S4). A hierarchical clustering analysis was applied to these probe sets and 138 breast cancers and 8 cell lines (Fig. 1A). The tumors displayed heterogeneous expression profiles. They were sorted into two large clusters, which nearly perfectly correlated with the molecular subtype, with all but one of the basal breast cancers in the left cluster and all but one of the luminal A breast cancers in the right cluster (Fig. 1B). Visual inspection revealed at least four clusters of related genes responsible for much of the subdivision of samples into two main groups. They are zoomed in Fig. 1C. The first cluster was enriched in genes involved in cell cycle and mitosis. It was overexpressed in basal overall compared with luminal A tumors and in cell lines compared with cancer tissue samples. The second gene cluster included many genes involved in immune reactions. It was expressed at heterogeneous levels in both luminal A and basal

tumors, and was overexpressed in lymphocytic cell lines compared with epithelial cell lines. The third and the fourth clusters were strongly overexpressed in luminal A overall compared with basal breast cancer samples. The third cluster included genes involved in transforming growth factor (TGF) $\beta$  signaling as well as transmembrane tyrosine kinase receptors. Gene ontology analysis using Ingenuity software confirmed these data with significant overrepresentation (right-tailed Fisher's exact test) of the functions "cell cycle" ( $P = 4.6E-07$ ) and "DNA replication, recombination, and repair" ( $P = 6.1E-05$ ) in the first cluster, "immune response" ( $P = 8.1E-10$ ) and "cellular growth and proliferation" ( $P = 8.1E-10$ ) in the second cluster, and "tumor morphology" ( $P = 2.2E-04$ ) and "nervous system development and function" ( $P = 2.3E-04$ ) in the third cluster. Analysis of canonical pathways showed overrepresentation of "G<sub>2</sub>-M transition of the cell cycle" ( $P = 6.8E-08$ ), "nuclear factor- $\kappa$ B signaling pathway" ( $P = 1.3E-04$ ), and "TGF $\beta$  signaling" ( $P = 4E-03$ ) in the first, second, and third clusters, respectively. No correlation was found between these gene clusters and the nine kinase families (*AGC*, *CAMK*, *CKI*, *CMGC*, *RGC*, *STE*, *TK*, *TKL*, and *Atypical*) or the chromosomal location of genes.

These results suggest that kinase gene expression is highly different between basal and luminal A breast cancers.



**Figure 1.** Kinase gene expression profiling in luminal A and basal breast cancers. **A**, hierarchical clustering of 138 breast cancer samples (80 luminal A and 58 basal; left), 8 cell lines (3 luminal epithelial mammary cell lines, 3 basal epithelial mammary cell lines, and 2 lymphocytic cell lines; right), and 435 unique kinase probe sets. Each row represents a gene and each column represents a sample. The expression level of each gene in a single sample is relative to its median abundance across the 138 breast cancer samples and is depicted according to a color scale shown at the bottom. In the right panel, genes are in the same order as in the left panel. Red and green, expression levels above and below the median, respectively. The magnitude of deviation from the median is represented by the color saturation. In the right panel, genes are in the same order as in the left panel. The dendrograms of samples (above matrix) represent overall similarities in gene expression profiles and are zoomed in **B**. Colored bars, the location of four gene clusters of interest (right) that are zoomed in **C**. **B**, dendrograms of samples. Top, dendrograms of breast cancer samples (left) and cell lines (right): two large groups of breast cancer samples are evidenced by clustering and delimited by dashed orange vertical line. Bottom, molecular subtype of samples (red, basal; blue, luminal A; green, lymphocytic cell lines). See the near-perfect separation of basal and luminal A breast cancers ( $P = 1.13 \times 10^{-36}$ , Fisher's exact test). **C**, expanded view of the four selected genes clusters. The first cluster is the 16 kinase gene cluster identified by quality threshold clustering. See its expression homogeneous in basal samples but rather heterogeneous in luminal A samples.

**Kinase gene expression identifies two subgroups of luminal A breast cancers.** As shown in Fig. 1, basal breast cancers constituted a rather homogenous cluster, whereas luminal A breast cancers were more heterogenous. Basal and luminal breast cancers were distinguished by the differential expression of clusters of genes. By using quality threshold clustering, we identified a single cluster of significance principally responsible for this discrimina-

tion (Fig. 1B), corresponding to the above-described first cluster. It contained 16 kinase genes (Table 1), which were overexpressed in all basal breast cancers and some luminal A samples, and underexpressed in most luminal A samples (Fig. 1B).

This subdivision of luminal A tumors led us to define for each of them the kinase score based on expression level of these 16 genes. A cutoff of 0 identified two tumor groups: a group containing the

**Table 1.** List of the 16 kinases from the gene cluster identified by quality threshold-clustering

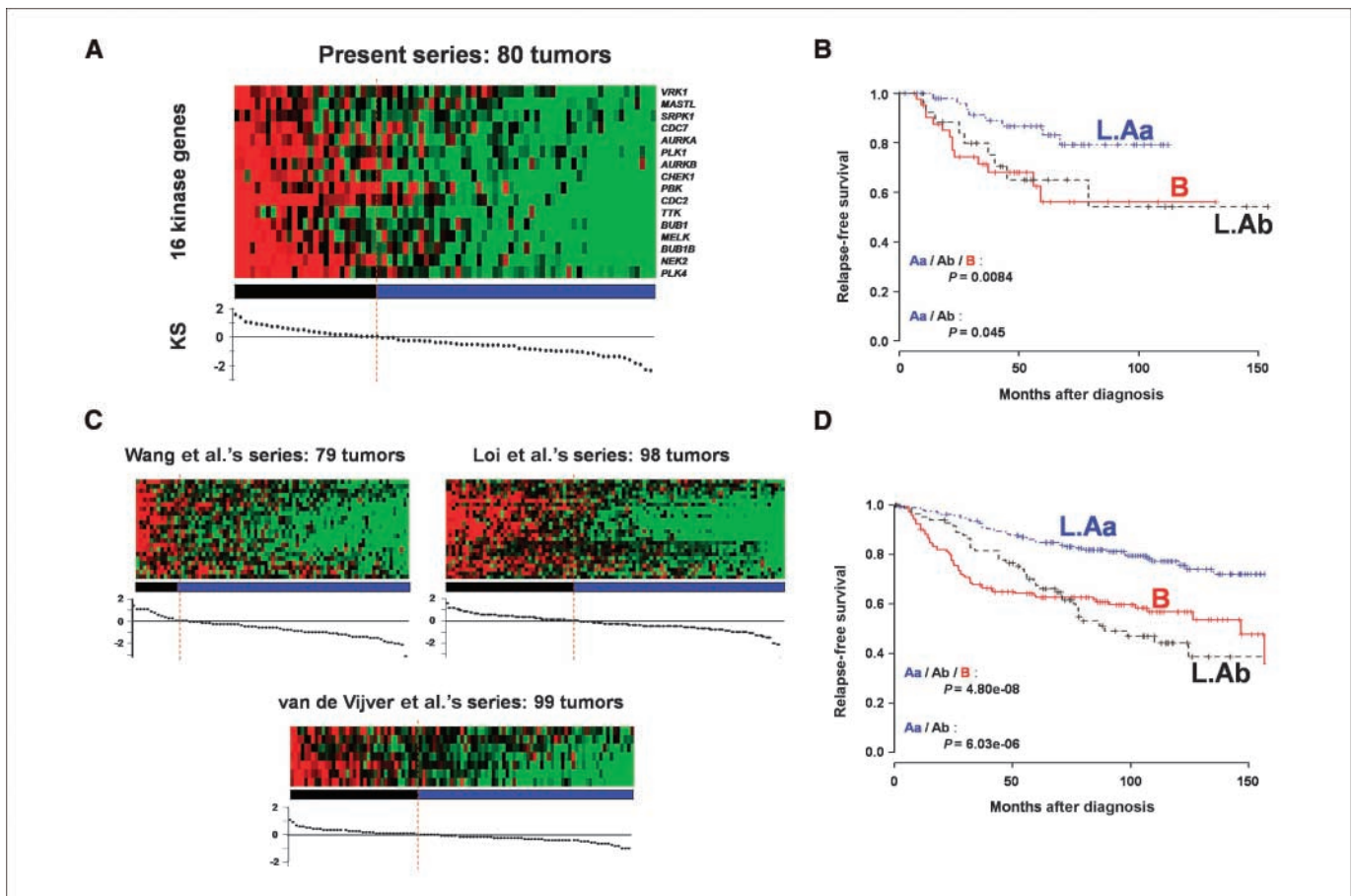
Probe set ID	Kinase activity	<i>P</i> *	Gene symbol	Names	Regulation	RefSeq transcript ID	Chrom. loc.	References for drugs
208079_s_at	Serine/threonine	2.06E-08	<i>AURKA</i>	Aurora kinase A, STK6, and STK15	Mitosis early phases and centrosome	NM_003600	20q13.2-q13.3	See Carvajal et al. (24)
209464_at	Serine/threonine	2.45E-13	<i>AURKB</i>	Aurora kinase B and STK12	Mitosis late phases and cytokinesis	NM_004217	17p13.1	See Carvajal et al. (24)
209642_at	Serine/threonine	3.84E-10	<i>BUB1</i>	Budding uninhibited by benzimidazoles 1 homologue (yeast)	Spindle assembly checkpoint	NM_004336	2q14	See de Carcer et al. (26)
203755_at	Serine/threonine	6.07E-12	<i>BUB1B</i>	Budding uninhibited by benzimidazoles 1 homologue $\beta$ (yeast) and BUBR1	Spindle assembly checkpoint	NM_001211	15q15	See de Carcer et al. (26)
203213_at	Serine/threonine	4.64E-16	<i>CDC2</i>	Cell division cycle 2, G <sub>1</sub> -S and G <sub>2</sub> -M, and CDK1	Cyclin complexes in G <sub>2</sub> -M	NM_001786	10q21.1	See de Carcer et al. (26)
204510_at	Serine/threonine	8.38E-06	<i>CDC7</i>	Cell division cycle 7 ( <i>S. cerevisiae</i> )	S phase prereplicative complexes	NM_003503	1p22	See de Carcer et al. (26)
205394_at	Serine/threonine	5.13E-10	<i>CHEK1</i>	CHK1 checkpoint homologue ( <i>S. pombe</i> )	S and G <sub>2</sub> phases, DNA damage checkpoint	NM_001274	11q24q24	See de Carcer et al. (26)
228468_at	Serine/threonine	8.65E-06	<i>MASTL</i>	Microtubule associated serine/threonine kinase like	Mitosis	NM_032844	10p12.1	
204825_at	Serine/threonine	2.30E-08	<i>MELK</i>	Maternal embryonic leucine zipper kinase and pEg3	G <sub>2</sub> -M transition and pre-mRNA splicing	NM_014791	9p13.2	
204641_at	Serine/threonine	6.85E-21	<i>NEK2</i>	Never in mitosis gene a-related kinase 2	Spindle assembly checkpoint and centrosome	NM_002497	1q32.2-q41	See de Carcer et al. (26)
219148_at	Serine/threonine	1.57E-10	<i>PBK</i>	PDZ binding kinase and TOBK	Mitosis	NM_018492	8p21.2	
202240_at	Serine/threonine	2.50E-13	<i>PLK1</i>	Polo-like kinase 1 ( <i>Drosophila</i> )	Spindle assembly checkpoint and centrosome	NM_005030	16p12.1	See Strebhardt and Ullrich (25)
204886_at	Serine/threonine	1.67E-08	<i>PLK4</i>	Polo-like kinase 4 ( <i>Drosophila</i> ) and SAK	Centrosome	NM_014264	4q27-q28	See Strebhardt and Ullrich (25)
202200_s_at	Serine/arginine	1.47E05	<i>SRPK1</i>	SFRS protein kinase 1	Pre-mRNA splicing	NM_003137	6p21.3-p21.2	
204822_at	Serine/threonine and tyrosine	5.88E-10	<i>TTK</i>	TTK (tramtrack) protein kinase and MPS1	Spindle assembly checkpoint	NM_003318	6q13-q21	See de Carcer et al. (26)
203856_at	Serine/threonine	2.05E-07	<i>VRK1</i>	Vaccinia-related kinase 1	S phase and P53 pathway	NM_003384	14q32	

NOTE: Variables for the quality threshold clustering was from 15 genes for minimum cluster size, with a minimum correlation of  $r = 0.70$ .

Abbreviation: Chrom. loc., chromosome location.

\**P* value for t test, to assume gene significance to separate both Luminal A groups.





**Figure 2.** Identification and validation of two prognostic subgroups of luminal A breast cancer samples based on the 16 kinase-gene set. **A**, classification of our 80 luminal A breast cancers using the 16 kinase genes. Genes are in the same order as in the cluster in Fig. 1C. Tumor samples are ordered from left to right according to the decreasing kinase score. Dashed orange line, the threshold 0 that separates the two classes of samples: luminal Ab with positive kinase score (at the left of the line; black horizontal class) and luminal Aa with negative kinase score (right to the line; blue horizontal class). Legend is as in Fig. 1. **B**, Kaplan-Meier RFS in our series of luminal Aa (L.Aa), luminal Ab (L.Ab), and basal (B) breast cancers. Basal medullary breast cancers were excluded from survival analyses. The  $P$  values are calculated using the log-rank test. **C**, classification of luminal A breast cancers from three public data sets using the 16 kinase genes: Wang et al. (15), Loi et al. (16), and van de Vijver et al. (14). The legend is similar to **A**. **D**, Kaplan-Meier RFS in the three pooled series of luminal Aa, luminal Ab, and basal breast cancers. The legend is similar to **B**.

luminal A breast cancers with negative score (hereafter designated Aa) and a group containing the luminal A breast cancers with positive score (hereafter designated Ab; Fig. 2A). Luminal Aa made up two-thirds of the luminal A cases, and luminal Ab breast cancers made up the remaining one-third.

Proteins encoded by the 16 genes overexpressed in luminal Ab breast cancers (Table 1) are all serine/threonine kinases (except SRPK1, which is a serine/arginine kinase) involved in the regulation of the late phases of the cell cycle, suggesting that luminal Ab tumors show a transcriptional program associated with mitosis.

**Characteristics and prognosis of the two subgroups of luminal A breast cancers.** The histoclinical characteristics of the two luminal A subgroups are listed in Table 2. Strikingly, they shared most features but were different according to Scarf-Bloom-Richardson (SBR) grade with more grade III in the Ab subgroup and more grade I to II in the Aa subgroup. Ki67 expression did not distinguish Ab from Aa cases but three-fourths of luminal Ab were Ki67-positive. In conclusion, no factor but grade could distinguish Aa from Ab breast cancers.

We compared the survival of three groups of patients, i.e., patients with basal, luminal Aa, and luminal Ab breast cancers. We

excluded from analysis the basal medullary breast cancers known to harbor good prognosis. With a median follow-up of 55 months after diagnosis, 5-year RFS (Fig. 2B) was best for patients with luminal Aa tumors (53 samples, 83% RFS) and worse for patients with luminal Ab tumors (27 samples, 65% RFS) and for patients with basal breast cancer (43 samples, 62% RFS;  $P = 0.031$ , log-rank test). Thus, the expression of 16 kinase genes identified within luminal A tumors of apparent good prognosis, a subgroup that showed a prognosis similar to basal cases.

We then compared the prognostic ability of our kinase score-based classifier with other histoclinical factors (age, pathologic tumor size, SBR grade, and axillary lymph node status, immunohistochemistry P53 and Ki67 status, and *ESR1* and *PGR* mRNA levels) in our 80 luminal A samples (Table 3A). In univariate and multivariate Cox analyses, the only factor that correlated with RFS was the kinase score-based classifier. The hazard ratio for relapse was 7.77 for luminal Ab tumors compared with luminal Aa tumors [95% confidence interval (CI) 1.97–30.66;  $P = 0.003$ ].

**Validation of two prognostic subgroups of luminal A breast cancers in published series.** As a validation step, we analyzed

**Table 2.** Histoclinical characteristics of the two luminal A tumor subgroups

Characteristics*		No. luminal A tumors (% of evaluated cases)			<i>P</i> <sup>†</sup>
		Total ( <i>n</i> = 80)	Luminal Aa subgroup ( <i>n</i> = 53)	Luminal Ab subgroup ( <i>n</i> = 27)	
Age (yr)	Median (range)	56 (24–82)	56 (28–82)	55 (24–82)	0.64
Pathologic type (80)	CAN	65 (81%)	41 (77%)	24 (89%)	0.28
	MIX	6 (8%)	5 (9%)	1 (4%)	
	LOB	9 (1%)	7 (14%)	2 (7%)	
Pathologic tumor size (69)	>2 cm	52 (66%)	34 (76%)	18 (75%)	1
	≤2 cm	17 (33%)	11 (24%)	6 (25%)	
SBR grade (79)	I–II	50 (63%)	41 (79%)	9 (33%)	1.50E–04
	III	29 (37%)	11 (21%)	18 (67%)	
Pathologic axillary lymph node status (76)	Positive	53 (66%)	35 (66%)	18 (66%)	0.8
	Negative	23 (33%)	14 (33%)	9 (33%)	
Immunohistochemistry ER status (80)	Positive	73 (91%)	46 (87%)	27 (100%)	0.089
	Negative	7 (9%)	7 (13%)	0 (0%)	
Immunohistochemistry PR status (80)	Positive	62 (78%)	39 (74%)	23 (85%)	0.27
	Negative	18 (22%)	14 (26%)	4 (15%)	
Immunohistochemistry P53 status (73)	Positive	15 (21%)	10 (22%)	5 (19%)	1
	Negative	58 (79%)	36 (78%)	22 (81%)	
Immunohistochemistry Ki67/MIB1 status (76)	Positive	47 (62%)	28 (57%)	19 (72%)	0.327
	Negative	29 (38%)	21 (43%)	8 (28%)	
Immunohistochemistry ERBB2 status (80)	Positive	4 (4%)	2 (4%)	3 (11%)	0.329
	Negative	76 (96%)	51 (96%)	24 (89%)	
<i>ESR1</i> mRNA level (80)	Rich	42 (53%)	25 (47%)	17 (63%)	0.238
	Poor	38 (47%)	28 (53%)	10 (37%)	
<i>PGR</i> mRNA level (80)	Rich	41 (51%)	26 (48%)	15 (56%)	0.641
	Poor	39 (49%)	27 (52%)	12 (44%)	
Relapse (80)	Yes	17 (21%)	8 (15%)	9 (33%)	0.083
	No	63 (79%)	45 (85%)	18 (67%)	
5-yr RFS (80)		76%	83%	65%	0.045

\*In parentheses are numbers of evaluated cases among 80 tumors.

†To assess differences in clinicopathologic features between the two groups of Luminal A patients, Fisher's Exact test was used for qualitative variables with discrete categories, the Wilcoxon test was used for continuous variables, and the log-rank test was used to compare Kaplan-Meier RFS.

three sets of published gene expression data to identify and compare the two subgroups of luminal A breast cancers identified by the kinase score. We first defined, as above, the molecular subtypes of tumors. Before assigning a subtype, each centroid was evaluated by its concordance with those defined by Sorlie et al. (4), and none was under 90% in the three data sets. The distribution of the subtypes is shown in Supplementary Table S5. A total of 276 samples were identified as luminal A. The number of genes in the 16 kinase gene set represented in each dataset ranged from 7 to 16 (Supplementary Table S5). We computed the kinase score for

each tumor. The same cutoff as in our series led to the identification of Aa (190 samples) and Ab (86 samples) subgroups in each set (Fig. 2C), with the same proportions as in our own series.

Samples from the three studies were pooled before prognostic analyses. Histoclinical correlations of the two subgroups were similar to those found in our series (Supplementary Table S6). We then compared RFS of the two luminal A subgroups in the 276 samples. With a median follow-up of 104 months after diagnosis, luminal Ab tumors were associated with a worse prognosis than luminal Aa tumors, with respective 5-year RFS of 90% and 73%

( $P = 6.3\text{E-}6$ , log-rank test; Fig. 2D). For comparison, 5-year RFS was 64% in basal samples in the three pooled series.

We also performed univariate and multivariate survival analyses (Table 3B). Wang et al's (15) series (79 Luminal A samples) was analyzed separately due to the lack of available histoclinical data. In univariate analysis, the hazard ratio for relapse was 4.84 for luminal Ab tumors compared with luminal Aa tumors (95% CI, 2.13–11.00;  $P = 1.7\text{E-}04$ ). The two other series were merged for analyses (197 Luminal A samples). Three variables, including pathologic tumor size, *PGR* mRNA expression level, and kinase score–based subgrouping, were significantly associated to RFS in univariate analysis. In multivariate analysis, only the kinase score–based classifier retained significant prognostic value, confirming

the prominence of the kinase score over the SBR grade and other variables. The hazard ratio for relapse was 2.48 for luminal Ab tumors compared with luminal Aa tumors (95% CI, 1.37–4.50;  $P = 0.002$ ).

**Kinase score and molecular subtypes.** We then studied the association of the kinase score with the intrinsic molecular subtypes. We merged all data sets, including our 227 tumors, the 295 van de Vijver et al's (14) tumors, the 414 Loi et al's (16) tumors, and the 286 Wang et al's (15) tumors, resulting in a total of 1,222 tumors. The kinase score and molecular subtypes were determined for all tumors: 367 tumors were luminal A, 99 luminal B, 172 ERBB2-overexpressing, 214 basal, 161 normal-like, and 209 unassigned. We computed and compared the distribution of the

**Table 3.**

**A. Univariate and multivariate RFS analyses by Cox regression of 80 luminal A tumors**

Variables	Univariate analysis				Multivariate analysis			
	<i>N</i> *	Hazard ratio	95% CI	<i>P</i>	<i>N</i> *	Hazard ratio	95% CI	<i>P</i>
This study								
Age >50 yr (vs ≤50 yr)	80	3.08	0.88–10.8	0.08	64	5.09	0.72–35.57	0.1
Pathologic tumor size >2cm (vs ≤2 cm)	69	1.9	0.54–6.75	0.32	64	4.77	0.86–26.41	0.07
SBR grade III (vs I + II)	79	1.71	0.66–4.46	0.27	64	1.62	0.43–6.03	0.47
Pathologic axillary lymph node status positive (vs negative)	80	1.57	0.51–4.82	0.43	64	1.43	0.32–6.24	0.63
Immunohistochemistry P53 status positive (vs negative)	73	1.65	0.52–5.27	0.4	64	1.62	0.37–7.01	0.52
Immunohistochemistry Ki67/MIB1 status positive (vs negative)	76	1.13	0.4–3.17	0.82	64	0.52	0.12–2.18	0.37
<i>ESR1</i> mRNA rich (vs poor)	80	2.09	0.73–5.94	0.17	64	1.12	0.2–6.27	0.9
<i>PGR</i> mRNA rich (vs poor)	80	0.64	0.24–1.68	0.36	64	0.23	0.05–1.06	0.06
KG subgroups LAb (vs LAa)	80	2.57	0.99–6.68	5.00E–02	64	7.77	1.97–30.66	3.40E–03

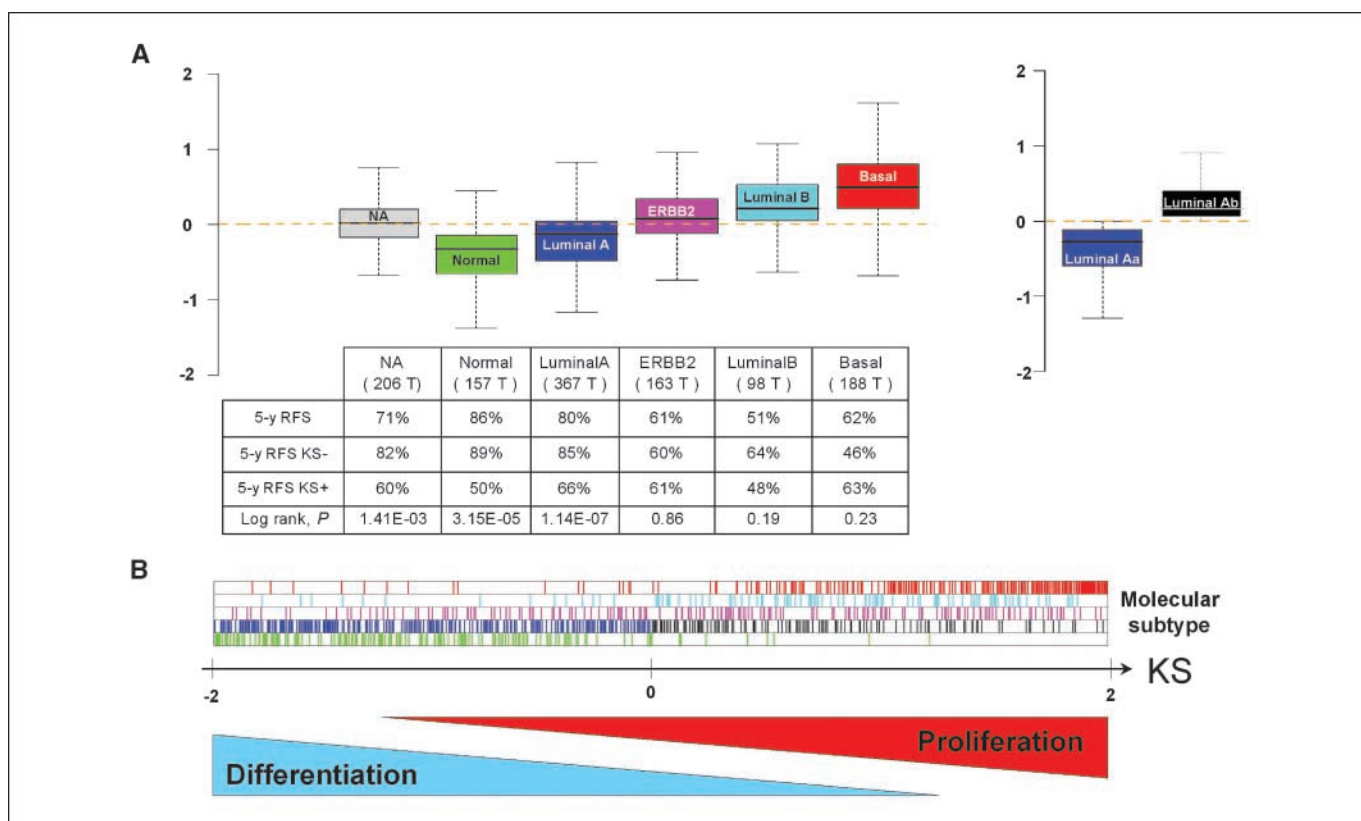
**B. Univariate and multivariate RFS analyses by Cox regression of luminal A tumors from published datasets**

Variables	Univariate analysis				Multivariate analysis			
	<i>N</i> *	Hazard ratio	95% CI	<i>P</i>	<i>N</i> *	Hazard ratio	95% CI	<i>P</i>
Loi and van de Vijver data sets								
Age >50 yr (vs ≤50 yr)	195	1.03	0.57–1.66	0.91	173	0.98	0.53–1.81	0.94
Pathologic tumor size >2cm (vs ≤2 cm)	195	2.04	1.19–3.5	9.80E–03	173	1.6	0.89–2.87	0.12
SBR grade III (vs I + II)	175	1.6	0.77–3.31	0.2	173	1.58	0.72–3.47	0.26
Pathologic axillary lymph node status positive (vs negative)	192	1.56	0.91–2.67	0.11	173	1.4	0.76–2.57	0.28
<i>ESR1</i> mRNA rich (vs poor)	195	0.67	0.38–1.18	0.17	173	0.8	0.42–1.51	0.49
<i>PGR</i> mRNA rich (vs poor)	195	0.44	0.26–0.76	3.00E–03	173	0.56	0.31–1.00	0.051
KG subgroups LAb (vs LAa)	195	3.07	1.78–5.29	5.50E–05	173	2.48	1.37–4.50	2.90E–03
Wang data set								
<i>ESR1</i> mRNA rich (vs poor)	79	0.75	0.35 to 1.61	0.47				
<i>PGR</i> mRNA rich (vs poor)	79	0.46	0.21 to 1.02	0.055				
KG subgroups LAb (vs LAa)	79	4.84	2.13 to 11.00	1.70E–04				

NOTE: Multivariate analysis not done for lack of annotations in Wang data set.

Abbreviation: KG, kinase gene.

\*Number of patients studied.



**Figure 3.** Kinase score in breast cancers. **A**, box plots of the kinase score (KS) in each molecular subtype (left) and each luminal A subgroup (right) across a total of 1,222 tumors. Median and range are indicated. NA, samples without any assigned subtype. Under the box plots are the 5-yr RFS for each subtype and for each kinase score-based subgroup in each subtype. Medullary breast cancers — all basal and one normal-like — were excluded from survival analyses. The *P* values are calculated using the log-rank test. **B**, classification of 1,222 tumors based on the kinase score. The molecular subtype of samples is indicated as follows: dark blue, luminal Aa; black, luminal Ab; light blue, luminal B; pink, ERBB2-overexpressing; red, basal; green, normal-like; and white, unassigned. Samples are ordered from left to right according to their increasing kinase score.

kinase score in each subtype. As shown in Fig. 3A, most of the luminal A and normal-like tumors had negative kinase score, whereas most of the basal and luminal B tumors had positive kinase score. All pairwise comparisons of kinase score between the five subtypes were significant ( $P < 0.05$ ; *t* test; data not shown). ERBB2-overexpressing and unassigned samples were equally distributed with respect to their kinase score. The luminal Ab tumors displayed a median kinase score, intermediate between that of luminal B tumors, to which the score was closer, and that of luminal Aa tumors.

The five molecular subtypes displayed different kinase score. However, because the range of kinase score was rather large in each subtype, we studied whether the kinase score had any prognostic value in other subtypes than luminal A by comparing survival (log-rank test) between kinase score-negative and kinase score-positive tumors (Fig. 3A). As expected, difference was strong in luminal A cases ( $P = 1.1E-07$ ). No difference was seen for ERBB2-overexpressing tumors ( $P = 0.86$ ). There was a nonsignificant trend ( $P = 0.18$ ) in luminal B tumors toward better RFS in kinase score-negative versus kinase score-positive samples. An opposite trend was observed in basal ( $P = 0.23$ ) with better RFS in kinase score-positive samples. The difference was strongly significant in normal-like tumors with 5-year RFS of 89% in kinase score-negative tumors and 50% in kinase score-positive tumors ( $P = 3.1E-05$ ). Interestingly, the kinase score could also be applied to the 209 samples not

assigned to a molecular subtype by the intrinsic gene set. It classified them in two prognostic subgroups, with difference for 5-year RFS between tumors with low kinase score (82%) and tumors with high kinase score (60%;  $P = 0.001$ ).

**A continuum in luminal breast cancers.** The luminal Ab tumors displayed an intermediate kinase score pattern between luminal Aa tumors and luminal B tumors (Fig. 3B). Comparison of histoclinical features between luminal Aa, luminal Ab, and luminal B samples in the three public data sets confirmed this finding (Supplementary Table S6), with a significant increase from luminal Aa to luminal Ab to luminal B for pathologic tumor size and rate of relapse, and a significant decrease for grade, mRNA expression level of *ESR1* and *PGR*, and 5-year RFS. These results confirm that luminal Aa and Ab represent new clinically relevant subgroups of breast cancers until now unrecognized and suggest a continuum between these three subgroups.

## Discussion

Our study focused on the kinome of luminal A and basal breast cancers, whose relevance to cancer biology and therapeutics is well-established (8). To our knowledge, this is the first study of profiling and exclusive and comprehensive analysis of kinase genes in breast cancer.

**The breast cancer kinome differs between luminal A and basal subtypes.** As an exploratory step, we applied hierarchical



clustering to 435 kinase genes. We found that luminal A and basal tumors had different global kinome expression patterns, with some degree of transcriptional heterogeneity within luminal A tumors. This observation suggests differential expression of many kinases and, consequently, different phosphorylation programs between the two subtypes. This result is not unexpected because kinases are involved in numerous pathways and many genes are differentially expressed between luminal A and basal breast cancers, which display numerous differentially activated functions.<sup>11</sup> Global clustering revealed broad coherent kinase clusters corresponding to cell processes (proliferation and differentiation) or to cell type (immune response), with overexpression of the proliferation cluster in basal samples and of the differentiation cluster in luminal A samples.

**Mitotic kinases identify two subgroups of luminal A breast cancers.** We identified a set of 16 genes sufficient to distinguish basal from luminal A tumors. Interestingly, a kinase score based on their expression distinguished two subgroups of luminal A tumors (Aa and Ab) with different survival. Identified in our tumor series, this classification and its prognostic effect were validated in 276 luminal A cases from three independent series profiled across different microarray platforms. Importantly, the kinase score outperformed the current prognostic factors in univariate and multivariate analyses in both training and validation sets.

Analysis of molecular function and biological processes revealed that the prognostic value of this kinase signature is mainly related to proliferation. Indeed, the 16 genes encode kinases involved in G<sub>2</sub> and M phases of the cell cycle. Aurora-A and Aurora-B are two major kinases regulating mitosis and cytokinesis, respectively. Budding inhibited by benzimidazole (BUB1), BUB1B, checkpoint kinase 1 (CHEK1), polo-like kinase (PLK)1, never in mitosis kinase 2 (NEK2), and TTK/MPS1 play key roles in the various cell division checkpoints. PLK4 is involved in centriole duplication. CDC2/CDK1 is a major component of the cell cycle machinery in association with mitotic cyclins. CDC7, maternal embryonic leucine zipper kinase (MELK), and vaccinia-related kinase 1 (VRK1) are regulators of the S-G<sub>2</sub> and G<sub>2</sub>-M transitions. SRPK1 regulates splicing. Not much is known about microtubule-associated serine/threonine kinase-like (MASTL) and PBK kinases.

Prognostic gene expression signatures related to grade (18, 19) or proliferation (20) have been reported. We found respectively 8 and 10 of our 16 kinase genes in the lists of genes differentially expressed in grade I versus grade III breast cancers reported by Sotiriou et al. (97 genes; ref 18) and Ivshina et al. (264 genes; ref. 19). Three kinase genes, *AURKA*, *AURKB*, and *BUB1*, are included in a prognostic set of 50 cell cycle-related genes (20), and *AURKB* is one of the five proliferation genes included in the Recurrence Score defined by Paik et al. (21). Furthermore, proliferation seems to be the most prominent predictor of outcome in many other published prognostic gene expression signatures (22). This link of our signature with proliferation also explains the correlation of our luminal A subgrouping with histologic grade, which is in part based on a mitotic index. But interestingly, comparison with Ki67 and grade showed that our mitotic kinase signature performed better in identifying these tumors and predicting the survival of patients.

**Mitotic kinases as therapeutic targets.** Targeting cell proliferation is a main objective of anticancer therapeutic strategies. Kinases have proven to be successful targets for therapies. Mitotic kinases have stimulated intense work focused on identifying novel antimitotic drugs. Some of them included in our signature represent targets under investigation (23). For example, targeting of Aurora kinases is a promising way of treating tumors (24). Clinical trials of four Aurora kinase inhibitors are ongoing in the United States and Europe: MK0457 and PHA-739358 inhibit Aurora-A and Aurora-B, MLN8054 selectively inhibits Aurora-A, and AZD1152 selectively inhibits Aurora-B. Similarly, small-molecule inhibitors of PLK1, such as ON01910 and BI2536, are being tested (25), as well as flavopiridol (inhibitor of the cyclin-dependant kinase CDC2) and UCN-01 (inhibitor of CHEK1). Other less studied but potential therapeutic targets include TTK, BUB, and NEK proteins (26).

**A new relevant subgroup of luminal A breast cancers.** Despite their relatively good prognosis compared with luminal B tumors, luminal A tumors display a heterogeneous clinical outcome after treatment, which generally includes hormone therapy. It is important to define the cases that may evolve unfavorably, all the more so that different types of hormone therapy, chemotherapy, and targeted molecular therapy are available. Our poor prognosis subgroup of luminal A tumors (Ab cases) is characterized by high mitotic activity compared with other luminal A tumors (Aa cases). Any error in the key steps in division regulated by these kinases — centrosome duplication, spindle checkpoint, microtubule-kinetochore attachment, chromosome condensation and segregation, and cytokinesis — may lead to aneuploidy and progressive chromosomal instability. This may in part explain the high grade and poor prognosis of these tumors.

In fact, the luminal Ab subgroup displayed clinical characteristics and a kinase score intermediate between the luminal Aa subgroup and the luminal B subtype. These subgroups were not previously recognized by the Sorlie's intrinsic gene set. We interpret this finding as follows. The use of intrinsic set distinguishes a large proportion of luminal B cancers but is unable to pick all proliferative cases. A small proportion of cases is left to cluster with the luminal A cases and are, therefore, labeled luminal A. An explanation for the poor efficacy of Sorlie's set to define all proliferative luminal cases may be the low number of genes involved in proliferation, including a very low number of kinases. Our mitotic kinase signature makes possible to identify all proliferative luminal cases and reveals a continuum of luminal cases from the more proliferative (luminal B) to the less proliferative (luminal Aa). Reciprocally, there may be a gradient of luminal differentiation giving a continuum of luminal breast cancers, including, from poorly differentiated to highly differentiated, luminal B, Ab, and Aa (Fig. 3B). Optimal response to hormone therapy would be obtained with luminal Aa breast cancers, whereas luminal B and Ab would benefit from chemotherapy and/or new drugs targeting the cell cycle and various kinases as discussed above.

## Acknowledgments

Received 9/18/2007; revised 11/9/2007; accepted 11/29/2007.

**Grant support:** Institut Paoli-Calmettes, Inserm, Université de la Méditerranée, and grants from the Institut National du Cancer (PL2006, ACI2007, and PHRC 2005 N°24-25), Association pour la Recherche sur le Cancer (2003 grant 3214), and Ligue Nationale contre le Cancer (label 2007).

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.

<sup>11</sup> Manuscript in preparation.

## References

1. Charafe-Jauffret E, Ginestier C, Monville F, et al. How to best classify breast cancer: conventional and novel classifications (review). *Int J Oncol* 2005;27:1307-13.
2. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747-52.
3. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869-74.
4. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418-23.
5. Bertucci F, Finetti P, Rougemont J, et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. *Cancer Res* 2005;65:2170-8.
6. Geyer CE, Forster J, Lindquist D, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* 2006;355:2733-43.
7. Hudis CA. Trastuzumab-mechanism of action and use in clinical practice. *N Engl J Med* 2007;357:39-51.
8. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* 2002;298:1912-34.
9. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Rev Cancer* 2004;4:177-83.
10. Krause DS, Van Etten RA. Tyrosine kinases as targets for cancer therapy. *N Engl J Med* 2005;353:172-87.
11. 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.
12. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249-64.
13. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863-8.
14. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999-2009.
15. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671-9.
16. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25:1239-46.
17. Bertucci F, Finetti P, Cervera N, et al. Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. *Cancer Res* 2006;66:4636-44.
18. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262-72.
19. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 2006;66:10292-301.
20. Dai H, van't Veer L, Lamb J, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 2005;65:4059-66.
21. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817-26.
22. Desmedt C, Sotiriou C. Proliferation: the most prominent predictor of clinical outcome in breast cancer. *Cell Cycle* 2006;5:2198-202.
23. Migliarese MR, Carlson RO. Development of new cancer therapeutic agents targeting mitosis. *Expert Opin Investig Drugs* 2006;15:1411-25.
24. Carvajal RD, Tse A, Schwartz GK. Aurora kinases: new targets for cancer therapy. *Clin Cancer Res* 2006;12:6869-75.
25. Strebhardt K, Ullrich A. Targeting polo-like kinase 1 for cancer therapy. *Nat Rev Cancer* 2006;6:321-30.
26. de Carcer G, de Castro IP, Malumbres M. Targeting cell cycle kinases for cancer therapy. *Curr Med Chem* 2007;14:969-85.