

# A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer

Philip Stephens<sup>1</sup>, Sarah Edkins<sup>1</sup>, Helen Davies<sup>1</sup>, Chris Greenman<sup>1</sup>, Charles Cox<sup>1</sup>, Chris Hunter<sup>1</sup>, Graham Bignell<sup>1</sup>, Jon Teague<sup>1</sup>, Raffaella Smith<sup>1</sup>, Claire Stevens<sup>1</sup>, Sarah O'Meara<sup>1</sup>, Adrian Parker<sup>1</sup>, Patrick Tarpey<sup>1</sup>, Tim Avis<sup>1</sup>, Andy Barthorpe<sup>1</sup>, Lisa Brackenbury<sup>1</sup>, Gemma Buck<sup>1</sup>, Adam Butler<sup>1</sup>, Jody Clements<sup>1</sup>, Jennifer Cole<sup>1</sup>, Ed Dicks<sup>1</sup>, Ken Edwards<sup>1</sup>, Simon Forbes<sup>1</sup>, Matthew Gorton<sup>1</sup>, Kristian Gray<sup>1</sup>, Kelly Halliday<sup>1</sup>, Rachel Harrison<sup>1</sup>, Katy Hills<sup>1</sup>, Jonathon Hinton<sup>1</sup>, David Jones<sup>1</sup>, Vivienne Kosmidou<sup>1</sup>, Ross Laman<sup>1</sup>, Richard Lugg<sup>1</sup>, Andrew Menzies<sup>1</sup>, Janet Perry<sup>1</sup>, Robert Petty<sup>1</sup>, Keiran Raine<sup>1</sup>, Rebecca Shepherd<sup>1</sup>, Alexandra Small<sup>1</sup>, Helen Solomon<sup>1</sup>, Yvonne Stephens<sup>1</sup>, Calli Tofts<sup>1</sup>, Jennifer Varian<sup>1</sup>, Anthony Webb<sup>1</sup>, Sofie West<sup>1</sup>, Sara Widaa<sup>1</sup>, Andrew Yates<sup>1</sup>, Francis Brasseur<sup>2</sup>, Colin S Cooper<sup>3</sup>, Adrienne M Flanagan<sup>4</sup>, Anthony Green<sup>5</sup>, Maggie Knowles<sup>6</sup>, Suet Y Leung<sup>7</sup>, Leendert H J Looijenga<sup>8</sup>, Bruce Malkowicz<sup>9</sup>, Marco A Pierotti<sup>10</sup>, Bin Teh<sup>11</sup>, Siu T Yuen<sup>7</sup>, Andrew G Nicholson<sup>12</sup>, Sunil Lakhani<sup>13</sup>, Douglas F Easton<sup>14</sup>, Barbara L Weber<sup>15</sup>, Michael R Stratton<sup>1,3</sup>, P Andrew Futreal<sup>1</sup> & Richard Wooster<sup>1</sup>

**We examined the coding sequence of 518 protein kinases, ~1.3 Mb of DNA per sample, in 25 breast cancers. In many tumors, we detected no somatic mutations. But a few had numerous somatic mutations with distinctive patterns indicative of either a mutator phenotype or a past exposure.**

We detected 92 somatic mutations, 90 base substitutions and two in-frame deletions (The Catalogue of Somatic Mutations in Cancer; **Supplementary Table 1** online) in the complete protein kinase gene family (**Supplementary Table 2** online)<sup>1</sup>. Of these, 58 base substitutions caused missense amino acid changes, 12 caused translational termination codons, 6 were at conserved positions in splice sites and 14 were silent (synonymous) changes.

The somatic mutations were distributed unevenly among the breast cancers that we examined. Twelve primary breast cancers had no somatic mutations; two had a single mutation each and one had two mutations. The remaining primary breast cancer (PD0119, an invasive pleomorphic lobular cancer in an 84-year-old woman, which we found by immunohistochemistry to be estrogen receptor-positive and E-cadherin-, *ERBB2*- and *TP53*-negative; **Supplementary Table 3** online) had 52 somatic mutations, all base substitutions. The mutations in this cancer had a distinctive pattern (**Fig. 1**). Mutations occurred with a high frequency at C:G base pairs (96%, 50 of 52), with many C:G→G:C transversions (44%, 22 of 50). Mutations occurred in a specific sequence context ( $P = 0.047$ ); specifically, at C:G base pairs, 94% (47 of 50) were 3' to a T:A base pair, compared with an expected frequency of 26% (**Fig. 1b**).

In contrast to the primary breast cancers, eight of the nine immortal breast cancer cell lines had at least one somatic mutation. The cell line with the largest number of somatic mutations, HCC2218 (derived from an invasive ductal carcinoma from a 38-year-old female, which was estrogen receptor-negative and *ERBB2*- and *TP53*-positive), had a mutational spectrum similar to that of PD0119 (**Fig. 1a**). All eleven of the somatic mutations detected in HCC2218 were at C:G base pairs; most were C:G→G:C transversions (56%, 6 of 11), and all arose at C:G base pairs 3' to T:A base pairs. The mutational spectra of HCC2218 and PD0119 differed from those of other cancers in the set ( $P = 0.003$  for heterogeneity) and from that of *TP53* obtained from several hundred breast cancers (**Fig. 1** and **Supplementary Tables 4** and **5** online). The results indicate that a previously hidden diversity of mutational processes exists in breast cancer.

The high frequency and distinctive pattern of somatic mutations observed in PD0119 and HCC2218 are probably due to a mutator phenotype (e.g., due to a DNA-repair defect), although mutagenic exposure cannot be excluded. *BRCA1* and *BRCA2* have been implicated in DNA repair. We did not find a truncating mutation in *BRCA1* or *BRCA2* in either PD0119 or HCC2218, but we did find one in a breast cancer cell line (HCC1395) that does not seem to have their characteristic mutational features. Neither PD0119 nor HCC2218 shows microsatellite instability (**Supplementary Methods** online). Therefore, it is unlikely that abnormalities in *BRCA1*, *BRCA2* or mismatch-repair genes account for the mutational spectra observed in PD0119 and HCC2218. These mutational spectra are also not readily attributable to other known DNA-repair defects. Yeast mutants

<sup>1</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. <sup>2</sup>Ludwig Institute for Cancer Research, 74 Avenue Hippocrate, B-1200 Brussels, Belgium. <sup>3</sup>Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK. <sup>4</sup>The Institute of Orthopaedics, University College London, Stanmore, Middlesex, UK. <sup>5</sup>Cambridge Institute for Medical Research, Addenbrooke's Hospital, Cambridge CB2 2XY, UK. <sup>6</sup>Cancer Research UK Clinical Centre, St. James University Hospital, Leeds, LS9 7TF, UK. <sup>7</sup>Department of Pathology, The University of Hong Kong, Hong Kong. <sup>8</sup>Josephine Nefkens Institute, Erasmus University, Rotterdam, 3015 GE Rotterdam, The Netherlands. <sup>9</sup>Division of Urology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>10</sup>Nazionale Tumori and FIRC Institute of Molecular Oncology, Milan, Italy. <sup>11</sup>Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA. <sup>12</sup>Royal Brompton Hospital, Sydney Street, London, SW3 6NP, UK. <sup>13</sup>Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, Chester Beatty Laboratories, London SW3 6JB, UK. <sup>14</sup>Cancer Research UK Genetic Epidemiology Group, Strangeways Research Laboratories, Cambridge, CB1 4RN, UK. <sup>15</sup>University of Pennsylvania Cancer Centre, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. Correspondence should be addressed to M.R.S. (mrs@sanger.ac.uk).

**Figure 1** Breast cancer mutational spectra.

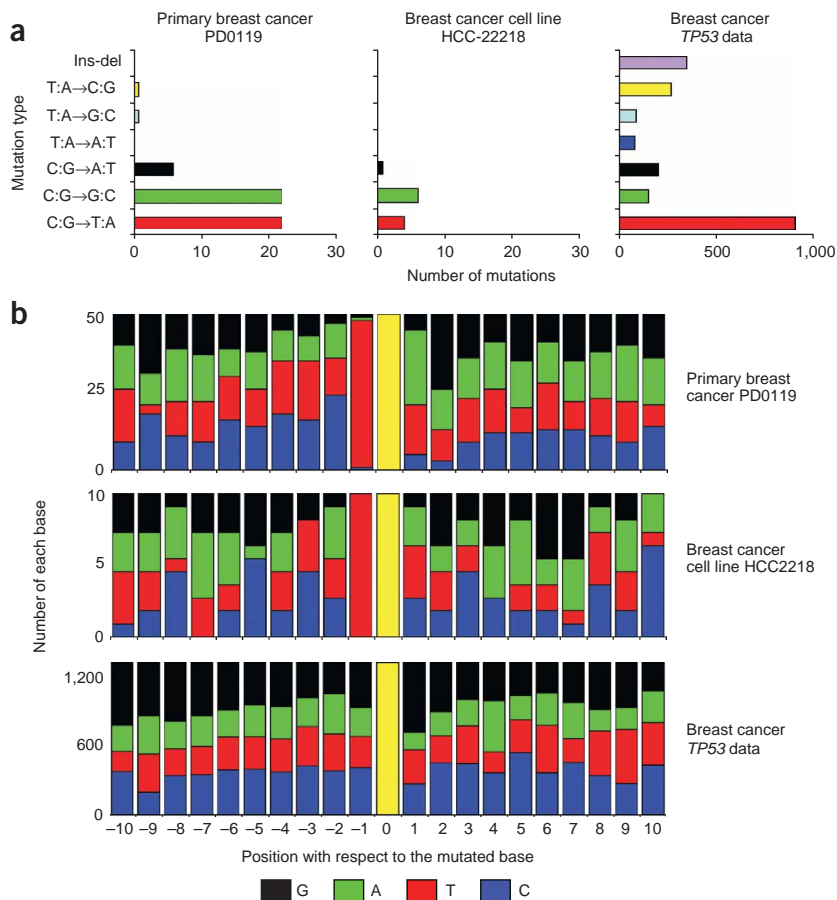
(a) Classes of mutation found in primary breast cancer PD0119 and in breast cancer cell line HCC2218 in the protein kinase genes and in *TP53* in several hundred breast cancers (International Agency for Research on Cancer TP53 mutation database, release R9)<sup>7</sup>.

(b) Sequence context of mutations at C:G base pairs found in PD0119 and HCC2218 in the protein kinase gene screen and in several hundred breast cancers analyzed through *TP53* (see also **Supplementary Tables 4 and 5** online). The yellow bar at position zero indicates the numbers of mutations at C:G base pairs. The remaining bars indicate the proportion of C (blue), G (black), A (green) and T (red) at positions 5' (minus) and 3' (plus) to the mutated C:G base pair.

exposed to various DNA-damaging agents show C:G→G:C transversion mutations but do not recapitulate the relative frequencies of mutations or sequence context that we observed<sup>2,3</sup>.

We next investigated whether some of the observed mutations are causally implicated in oncogenesis. The 92 somatic mutations were distributed among 73 genes (of which 65 had nonsynonymous or splice-site changes). Monte-Carlo simulations (accounting for transcript sizes) showed that this distribution does not deviate from that expected by chance. Moreover, there was no clustering of missense mutations within kinase domains (as might be expected for activating mutations) or in tyrosine, serine-threonine or atypical kinases.

Mutations that contribute to oncogenesis are likely to change the amino acid sequence. Therefore, a smaller proportion of observed synonymous mutations than expected would be evidence of biological selection. Seventy-six (84%) of the 90 observed base-substitution mutations cause nonsynonymous changes, compared with 69% expected on the basis of the composition of sequence examined and mutation spectrum observed. The distribution of mutation types differed significantly from that expected by chance (**Table 1**;  $P = 0.0002$ ). Approximately 38% more nonsynonymous mutations occurred than would be expected by chance, predominantly composed of nonsense changes (rate ratio



4.0,  $P = 0.001$ ) and splice-site variants (rate ratio 4.5,  $P = 0.006$ ). There was a slight, though nonsignificant, excess of missense variants (rate ratio 1.4, 95% confidence interval = 0.8–2.4). These excesses were not attributable to PD0119 alone. Overall, the analyses indicate that a subset of the somatic mutations detected contributes to cancer development.

All nonsense and five of six splice-site variants were present in a heterozygous state. Ten of twelve nonsense mutations and four of six splice-site variants were in PD0119, the primary breast cancer with a putative mutator phenotype. Eight of nine nonsense changes that could be evaluated were present in transcripts and hence were not subject to nonsense-mediated RNA decay<sup>4</sup>. In general, truncating mutations observed in classical tumor-suppressor (recessive cancer)

**Table 1** Rates of missense, nonsense and splice mutation types

Sample set	All samples			PD0119			Remaining samples		
	Observed	Expected	Rate ratio	Observed	Expected	Rate ratio	Observed	Expected	Rate ratio
Synonymous	14	20.7	1	10	12.3	1	4	8.5	1
Nonsynonymous									
Total	76	69.3	1.6	42	39.7	1.3	34	29.6	2.4
Missense	58	62.9	1.4	28	35.8	1	30	27.1	2.3
Nonsense	12	4.4	4	10	2.9	4.2	2	1.5	2.8
Splice	6	2	4.5	4	1	4.8	2	0.9	4.5

The expected rates were calculated from the number of mutable base pairs screened, while preserving the observed mutation spectra, under the assumption of no selection pressure. The rate ratio divides the observed and expected rates; values greater than 1 indicate positive selection.

genes affect both alleles. Therefore, the process by which these heterozygous truncating mutations contribute to oncogenesis is uncertain.

We sequenced the 65 genes in which we found nonsynonymous somatic mutations in an additional set of 56 primary breast cancers. We observed three additional nonsynonymous mutations (including one frameshift mutation) and one silent mutation in four genes. Neither the overall mutation rate nor the ratio of nonsynonymous to synonymous mutations differed significantly from that predicted from the primary screen.

This is the first mutational screen to our knowledge of the full coding sequence of all protein kinases in cancer. The results suggest that there is no commonly point-mutated and activated protein kinase gene in invasive ductal breast cancer despite amplification of the receptor tyrosine kinase ERBB2 in 20% of cases<sup>5</sup>. Larger series will be necessary to evaluate the presence of relatively frequently mutated protein kinases in subclasses of breast cancer. Moreover, we cannot exclude the possibility that single missense mutations of the genes *LYN* or *MAST2* (in paralogous positions to activating, oncogenic mutations in other kinases) or translational termination mutations in the gene *ROCK1* (found in both the primary and follow-up series) were implicated in the development of the cancers in which they were found.

More than 3% of the coding sequence of the human genome has been analyzed in each breast cancer. This provides an opportunity to examine the mutational landscape of breast cancer genomes. On the basis of our analysis, approximately two-thirds of the observed mutations are probably 'passenger' mutations that are not subject to selection and, therefore, occur at a similar frequency throughout the

genome. With this assumption, the results indicate that many primary breast cancers typically have 3,000 or fewer somatically acquired point mutations in their genomes (approximately one or fewer per megabase) and hence fewer than 30 randomly generated, 'passenger' amino acid changes. These results are consistent with previously published observations in colorectal cancer, which were based on ~10% of the amount of DNA that we screened<sup>6</sup>. But a few breast cancers have many more mutations. We estimate that PD0119 has ~100,000 somatic mutations in its genome (30 per megabase), which would result in ~1,000 amino acid changes (one in every 20–30 genes).

**URLs.** The Catalogue of Somatic Mutations in Cancer is available at <http://www.sanger.ac.uk/cosmic>. The International Agency for Research on Cancer TP53 mutation database is available at <http://www-p53.iarc.fr/index.html>.

*Note: Supplementary information is available on the Nature Genetics website.*

#### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 23 December 2004; accepted 15 April 2005

Published online at <http://www.nature.com/naturegenetics/>

1. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. *Science* **298**, 1912–1934 (2002).
2. Huang, M.E., Rio, A.G., Nicolas, A. & Kolodner, R.D. *Proc. Natl. Acad. Sci. USA* **100**, 11529–11534 (2003).
3. Xiao, W., Fontanie, T., Bawa, S. & Kohalmi, L. *Mutat. Res.* **431**, 155–165 (1999).
4. Wagner, E. & Lykke-Andersen, J. *J. Cell Sci.* **115**, 3033–3038 (2002).
5. Persons, D.L., Borelli, K.A. & Hsu, P.H. *Mod. Pathol.* **10**, 720–727 (1997).
6. Wang, T.L. *et al. Proc. Natl. Acad. Sci. USA* **99**, 3076–3080 (2002).
7. Olivier, M. *et al. Hum. Mutat.* **19**, 607–614 (2002).