Oncogene (2005) 24, 5235–5245

www.nature.com/onc

A 1 Mb minimal amplicon at 8p11–12 in breast cancer identifies new candidate oncogenes

Maria J Garcia[1], Jessica CM Pole[2], Suet-Feung Chin[1], Andrew Teschendorff[1], Ali Naderi[1], Hilal Ozdag[1], Maria Vias[1], Tanja Kranjac[1], Tatiana Subkhankulova[1], Claire Paish[3], Ian Ellis[3], James D Brenton[1], Paul AW Edwards[2] and Carlos Caldas*,[1]

[1]*Department of Oncology, Hutchison/MRC Research Centre, Cancer Genomics Program, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK;* [2]*Department of Pathology, Hutchison/MRC Research Centre, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK;* [3]*Department of Histopathology, The Breast Unit, Nottingham City Hospital NHS Trust and University of Nottingham, Nottingham NG5 1PB, UK*

Amplification of 8p11–12 is a well-known alteration in human breast cancers but the driving oncogene has not been identified. We have developed a high-resolution comparative genomic hybridization array covering 8p11–12 and analysed 33 primary breast tumors, 20 primary ovarian tumors and 27 breast cancer cell lines. Expression analysis of the genes in the region was carried out by using real-time quantitative PCR and/or oligo-microarray profiling. In all, 24% (8/33) of the breast tumors, 5% (1/20) of the ovary tumors and 15% (4/27) of the cell lines showed 8p11–12 amplification. We identified a 1 Mb segment of common amplification that excludes previously proposed candidate genes. Some of the amplified genes did not show overexpression, whereas for others, overexpression was not specifically attributable to amplification. The genes *FLJ14299*, *C8orf2*, *BRF2* and *RAB11FIP*, map within the 8p11–12 minimal amplicon, two have a putative function consistent with an oncogenic role, these four genes showed a strong correlation between amplification and overexpression and are therefore the best candidate driver oncogenes at 8p12.
Oncogene (2005) **24,** 5235–5245. doi:10.1038/sj.onc.1208741; published online 16 May 2005

Keywords: 8p11–12 amplicon; breast cancer; array-CGH; minimal region of amplification; candidate oncogene

Introduction

The short arm of chromosome 8 is one of the most frequently altered regions in human cancers. Loss of heterozygosity (LOH) on 8p has been described in most of the common epithelial tumors such as colorectal, lung, prostate, ovary or breast carcinoma and different candidate tumor suppressor genes have been proposed (Emi *et al*., 1992; Bova *et al*., 1996; Pribill *et al*., 2001; Toomes *et al*., 2003; Bhattacharya *et al*., 2004; Flanagan *et al*., 2004). Chromosomal translocations in 8p are commonly found in hematologic malignancies such as acute myeloid leukemia, and are also reported in breast tumors (Borrow *et al*., 1996; Wang *et al*., 1999; Adelaide *et al*., 2003; Huang *et al*., 2004). Amplification of 8p11–12 is a well documented event in breast tumors, but the relevant oncogene for the region has not been clearly elucidated (Theillet *et al*., 1993; Ugolini *et al*., 1999). The *FGFR1* gene, which belongs to the fibroblast growth factor receptor family, has been long considered a strong candidate oncogene for the amplicon (Theillet *et al*., 1993; Adelaide *et al*., 1998; Ugolini *et al*., 1999). However, functional analysis has failed to provide direct evidence of an oncogenic role for *FGFR1* and its importance as a driver of the amplicon has been questioned (Ray *et al*., 2004).

The use of incomplete maps and analysis of different markers has hampered definition of the key oncogene in the 8p11–12 region. With the release of the Human Genome Sequence and its subsequent refined versions very accurately annotated clone and gene assemblies for this region are available. In addition, the development of array comparative genomic hybridization (array-CGH) allows reliable assessment of DNA copy-number changes in a high-throughput manner and has proved to be very useful in the characterization of well-known amplicons (Pinkel *et al*., 1998; Albertson *et al*., 2000; Garnis *et al*., 2004). In a recent study, 1 Mb coverage array-CGH in combination with Southern blot analysis was used to characterize the 8p11–12 amplicon in three breast cell lines (Ray *et al*., 2004). However, very high-resolution analysis of 8p11–12 has not been performed, which prompted us to develop a near-tiling path CGH-array for this region. We report here a detailed genomic and expression analysis of the 8p11–12 region in a panel of samples comprising breast cancer cell lines and primary breast and ovarian tumors.

Results

Whole genome overview

High-resolution analysis of chromosome 8 was provided by clones specifically selected for the 8p11–12 region in a

*Correspondence: C Caldas; E-mail: cc234@cam.ac.uk
Received 21 January 2005; revised 22 March 2005; accepted 6 April 2005; published online 16 May 2005

near-tiling path and elsewhere on chromosome 8 at approximately 1.5 Mb resolution. Clones covering the whole genome at approximately 10 Mb intervals allowed us to identify genome-wide copy-number changes at low resolution and hence gain overall impression of performance of the array (Figure 1 and Supplementary Table 4). Briefly, the most frequently gained regions in the primary breast tumors and breast cell lines were 8q, 1q, 20q, 7p and 12p, and regions of amplification were 8p11–12, 17q12, 8q24, 11q13 and 20q13. Common regions of copy loss were 17p, 22q, 9q, 8p, 1p and 18q. In the ovarian tumors, the most recurrently gained regions were 8q, 20q, 3q, 9p, 1q and 12p, and regions of amplification were 8q24, 15q21, 3q26 and 8p12. The most common regions of loss were 22q, 15q, 17p, 9q or 17q.

Fine mapping of the 8p11–12 amplicon

A total number of 13 tumors and cell lines showed amplification of 8p11–12, defined as a normalized $\log_2$ ratio greater than 1.0. Most of them were breast malignancies: eight out of 33 primary breast tumors (24%) and four out of 27 (15%) breast cancer cell lines. Only one primary ovarian tumor out of 20 analysed exhibited amplification at 8p11–12 (5%). The size of the amplicon was variable and ranged from as large as 11 Mb (SUM52) to as small as 1 Mb (tumor 2156) (Table 1, Figure 2). The SUM52 amplicon, which was the largest one, spanned 31.4 to 42.5 Mb, but for the majority of samples, the amplicon boundaries were within the region between 35 and 40 Mb. However, the precise flanking sites varied across samples. Remarkably, a steep copy-number transition delimiting the distal end of the aberration was found at the same position in four breast tumors (35.5 Mb, clones RP11–20J10 and RP11–115J10; tumors 2105, 2199, 2219 and 2218). The only gene mapping at this location is *UNC5D*, a member of the UNC5 family of netrin-1 receptors. In none of the cases with 8p11–12 amplification did the amplicon extend across the centromere (Figure 2). In a few cases (tumors 2017, 2219, 2156 and 8-TB010321), 8q11 was gained but as a consequence of the complete or partial gain of 8q and independent of the amplification at 8p12.

A minimal common region of amplification of 1 Mb was defined by tumors 2156, 8-TB010321 and the breast cell line HCC1500 (Figures 2 and 3a). The region was bounded by the nonamplified clones RP11–175F15 (36.9 Mb) and RP11–701H6 (37.9 Mb). The genes within this region according to NCBI Build 35 of the Human Genome Sequence are only six: *FLJ14299*, *C8orf2*, *PROSC*, *GPR124*, *BRF2* and *RAB11FIP1* (Figure 2). Importantly, other previously proposed candidate genes such as *LSM1*, *BAG4*, *HTPAP*, *FGFR1*, *TACC1* or *SFRP1* map beyond the boundaries of this segment. All the cases presented slightly varying copy number across the region, but most of them did not show abrupt

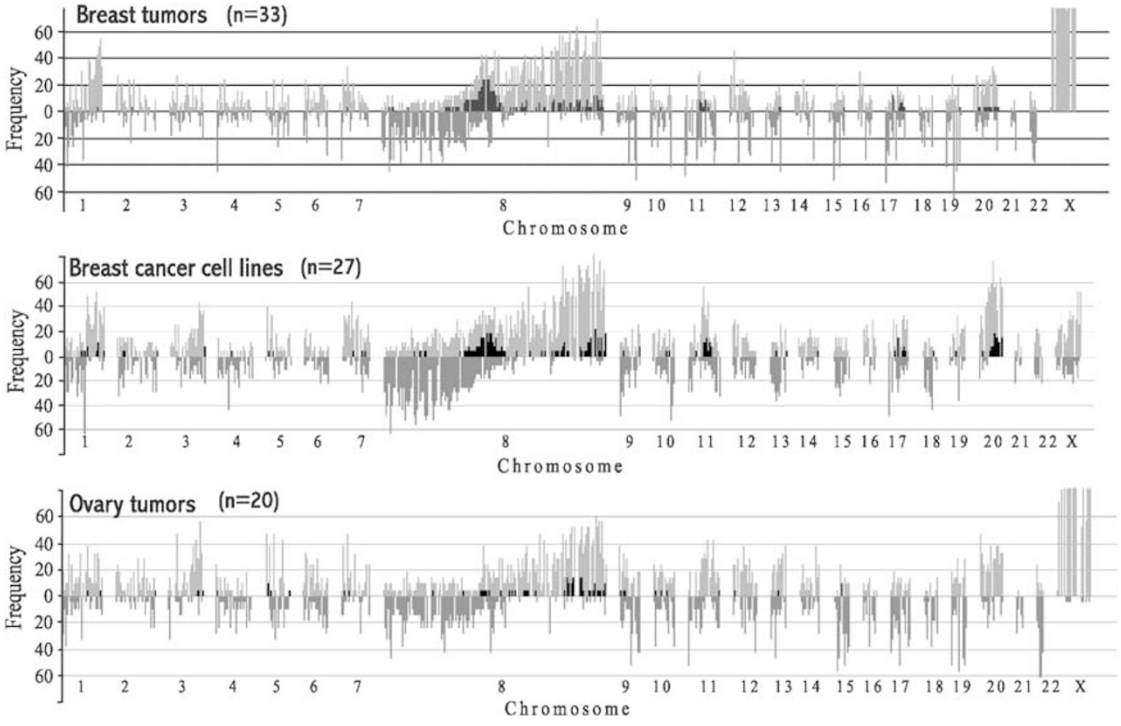


Figure 1 Genome-wide frequency of copy-number changes in breast tumors, breast cancer cell lines and ovarian tumors. Green bars represent the percentage of cases with gain for each clone. Red bars represent the percentage of cases with loss for each clone. Black bars represent the percentage of cases with amplification. The thresholds for gain, loss and amplification were set at ratios >1.2 ($\log_2$ ratio >0.26), ratios <0.8 ($\log_2$ ratio <-0.32) and ratios >2 ($\log_2$ ratio >1), respectively. Clones are ordered according to their position along the chromosomes (NCBI Build 35) in the *X*-axis. The average distance between clones is 10 Mb, except for chr. 8, where it is 1.5 Mb increased to near-tiling path resolution at 8p11–12

Table 1 Amplicon boundaries, size and structure

Amplified cases	*Start*[a] *(Mb)*	*End*[a] *(Mb)*	*Size*[b] *(Mb)*	*Peaks of amplification*[c] *(log_2 ratios)*				
				34 Mb	*35.6 Mb*	*37 Mb*	*38 Mb*	*40.5 Mb*
SUM52	31.38	42.58	11.2	[d]	[d]	[d]	[d]	[d]
MDA-MB-134	34.67	42.58	7.91		3	3.13		2.5
2105	35.49	43.38	7.89			2.3		
2107	36.09	43.38	7.29			1.8		
SUM44	36.01	42.58	6.57			2.8		2.1
2199	35.13	41.52	6.39			1.8		1.65
2171	36.62	42.58	5.96	1.34		1.2		2.13
2219	35.13	40.25	5.12			1.8		
2218	35.13	40.25	5.12		2.1	2.3		
2138	36.62	39.40	3.63			2.7	3.7	
HCC1500	36.09	38.16	2.07			2.8		
8-TB010321	36.44	37.99	1.55			1.8		
2156	36.95	37.99	1.04			1.8		

[a]Midpoints of the clones flanking the amplicon. In case 2171, the start position does not take into account the peak at 34 Mb, which is not contiguous to the rest of the amplicon. [b]Calculated as the distance between the midpoints of the clones flanking the amplicon. [c]Peaks of amplification within the amplicon. The highest log_2 ratio for each core is shown. [d]SUM52 did not show peaks of amplification but a continuous log_2 value of 1

changes indicative of several cores of amplification. There were nevertheless some exceptions that showed a more complex structure. The profile of MDA-MB-134 appeared to have three peaks of amplification, one centered at the region of common amplification and two additional peaks centered at positions 35.6 Mb (RP11–115J10) and 40.5 Mb (RP11–51K1237) (Figure 3b and Table 1). SUM44 and the breast tumors 2199 and 2171 also presented the peak at 40.5 Mb. Tumor 2171 showed an extra site of amplification located around 34 Mb that was not observed in any of the other samples (Figure 3b). Tumor 2138 presented two distinct levels of amplification consisting of a plateau at the minimal region of amplification and an adjacent higher peak at 38.4 Mb (Figure 3b). This peak was unique to this sample. There were no remarkable differences between profiles obtained in the primary tumors and cell lines. The general higher level of amplification in cell lines is probably due to the presence of contaminating normal cells in tumors that result in suppression of ratio changes.

In addition to the amplification, we also found two small regions of loss (Table 2). Copy-number transitions that suggested breakpoints of unbalanced translocations within 8p11–12 were observed in four breast tumors, but did not have a common location. In two cases clones spanning the 3′-end of the *NRG1* and *UNC5D* genes, respectively, flanked the breaks. Another region of chromosome 8 showing copy-number changes was 8p23, which was frequently lost in both primary tumors and cell lines (Table 2).

In order to validate the results of the array-CGH, we performed interphase fluorescence in situ hybridization (FISH) on imprints from frozen tumors. We chose tumors showing marked differences in their profile and a set of probes adequate to test that variation. Despite the limitations imposed by working with primary tumors in array-CGH, such as the presence of contaminating normal cells and unknown ploidy, the copy-number changes estimated by array-CGH were in very good agreement with those shown by the FISH experiments (Figure 2b and c). We also verified by metaphase FISH the results obtained for HCC1500 (Figure 2d), which showed the smallest amplicon within the cell lines.

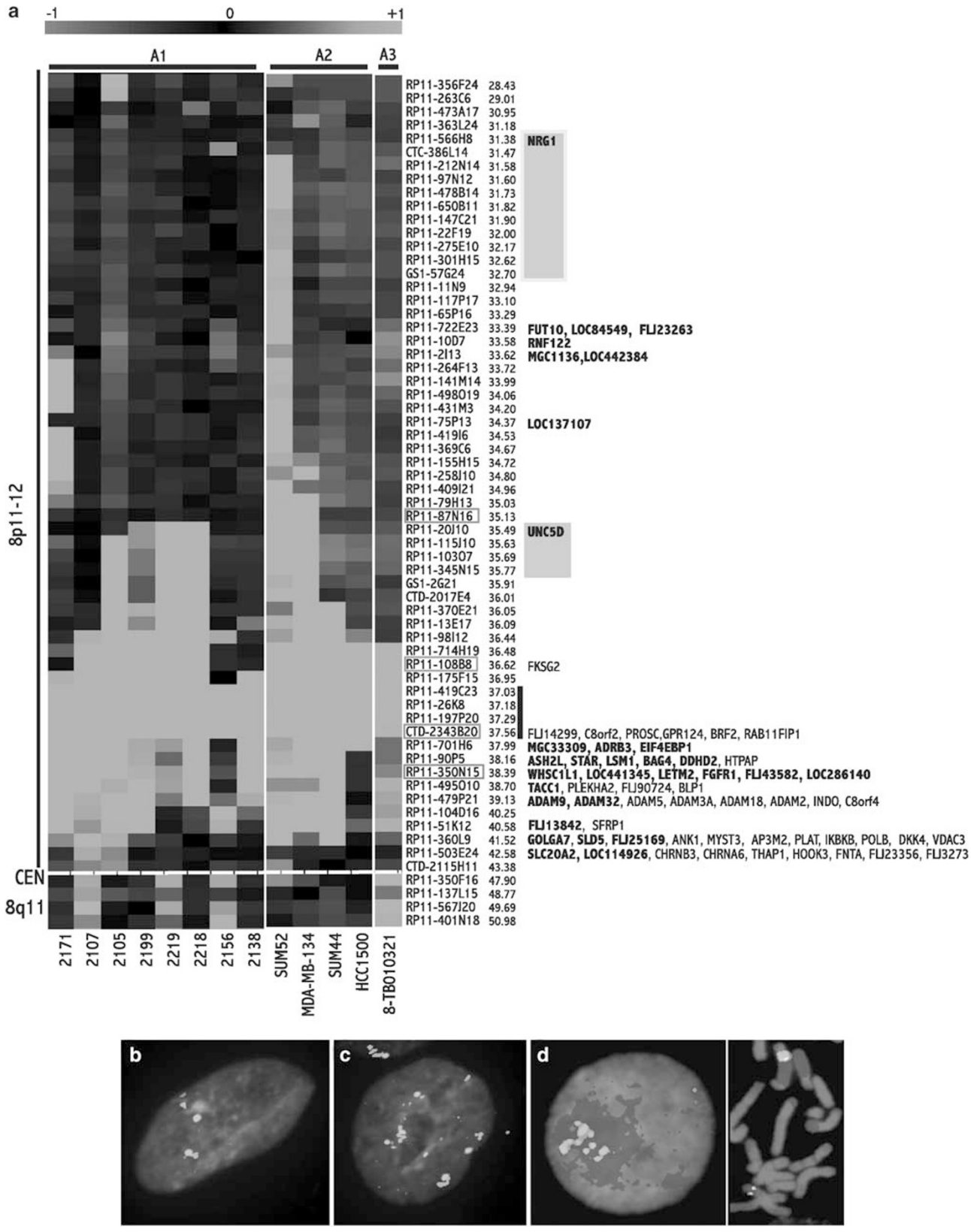
All eight cancers with amplification were histological grade 2 or 3 and five (62%) developed distant metastasis during follow-up. In the group of nonamplified tumors, 24% were grade 1 and 28% developed distant metastasis. Neither these nor other clinical variables such as stage, disease recurrence, overall survival and estrogen receptor status were significantly associated with the presence of 8p11–12 amplification in the series of 33 breast tumors analysed (Fisher's test, $P > 0.1$). Similarly, the Nottingham Prognostic Index (Haybittle *et al*., 1982; Todd *et al*., 1987) treated as a continuous variable was not significantly associated with amplification (Pearson's correlation value of 0.13; Monte-Carlo simulation, $P > 0.1$). However, it is likely that additional samples are required to fully explore the relationship between amplification at 8p11–12 and clinical parameters.

Expression analysis

In order to determine whether 8p11–12 amplification was associated with gene overexpression we performed real-time quantitative PCR (rtq-PCR) analysis of 10 genes in the region, three located within the minimal amplicon. We studied 51 samples previously analysed by array CGH (26 primary breast tumors and 27 breast cell lines). Except for two tumors (2218 and 2219), we found that all the amplified samples showed overexpression relative to normal luminal breast cells for most of the genes within the amplicon (Figure 4a). Among the three genes mapping within the minimal region of amplification, *GPR124* and *FLJ14299* demonstrated high levels of expression with a median expression increase in the amplified samples relative to normal luminal cells of

13- and four-fold, respectively. *RAB11FIP1* was also overexpressed but at a more moderate level, with a median ratio of 1.6.

To assess whether gene overexpression was specifically associated with amplification, we compared gene expression in the amplified samples with that observed



in a set of samples where we did not detect any copy-number changes at 8p11–12. The level of expression in the two groups (Figure 4b) was significantly different for *FLJ14299*, *RAB11FIP1*, *LSM1*, *BAG4* and *HTPAP* (Wilcoxon's rank-sum test, $P<0.05$). Significantly, *GPR124* (located within the minimal amplicon) and *FGFR1* overexpression was not specifically related to the presence of amplification. For example, *FGFR1* was overexpressed in cell line HCC1500, where we verified by FISH that *FGFR1* was not part of the amplicon (Figure 2d), which suggests that other mechanisms are altering *FGFR1* expression. *UNC5D* did not show a pattern of expression that correlated with DNA copy-number changes. The gene was under-expressed compared to normal luminal cells in most of the samples, including those where it was part of the amplicon. Furthermore, among samples whose expression was higher, some showed copy-number loss in the region.

In all, 29 out of the 33 primary breast tumors analysed for DNA copy-number changes, including seven out of the eight amplified cases, were also characterized by expression profiling using a human oligo microarray (Figure 4c). We retrieved the expression microarray data for 31 genes included in the array mapping to 8p11–12. This allowed us to complete the expression analysis in the minimal region of amplification and get additional information on flanking genes and genes with a more centromeric location. Seven genes were analysed using both rtq-PCR and oligo-microarray profiling (Figure 4a and c), and the correlation of expression values was good for five of the genes (*FLJ14299*, *RAB11FIP1*, *LSM1*, *BAG4* and *HTPAP*; Pearson's coefficients >0.55; Mantel test, $P<0.05$). As

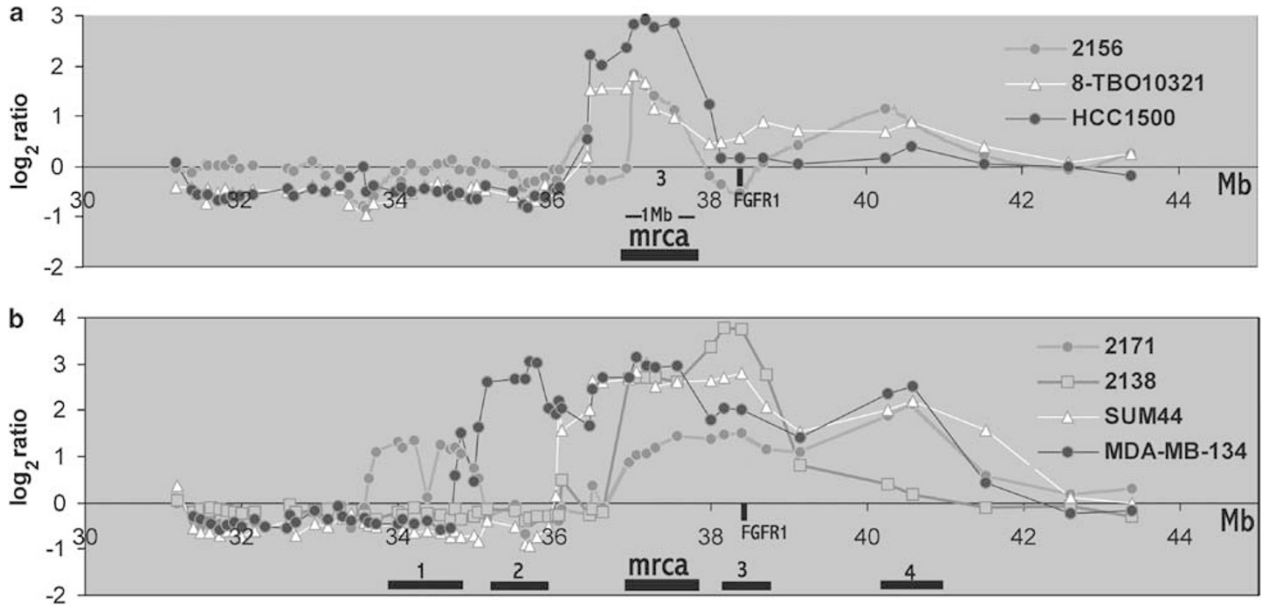


Figure 3 Genomic copy-number profiles of 8p11–12 in selected cases. (**a**) Profiles of cases with narrow regions of copy-number maxima that define a minimal region of common amplification (black bar). (**b**) Profiles of cases that showed a complex amplicon structure. The minimal common region of amplification and possible cores of independent amplification are represented by black bars (labeled mrca and 1–4). Cores 1 and 3 were unique to samples 2171 and 2138, respectively. Log_2 ratios are plotted against the clone position according to the NCBI Build 35 of the Human Genome sequence

Figure 2 (**a**) Array-CGH across the 8p11–12 region. The eight breast tumors (A1), four breast cell lines (A2) and single ovarian tumor (A3) that showed amplification at 8p11–12 are represented in columns. Each row represents a clone on the array. Clones have been ordered by genome position according to the NCBI Build 35 from the most distal ones (top) to the most centromeric ones (bottom). Framed clones were used in the FISH validation experiments. Clones covering 8q11 have also been included. The midpoint position (Mb) of each clone is indicated beside the clone name as well as the genes mapping in the region. Genes in bold font are totally or partially contained within the clone. Genes in regular font map to gaps between clones and have been assigned to the clone immediately distal to the gap region. Log_2 ratios are depicted in a color scale where green represents gain and red loss. Grey cells correspond to data rejected after quality tests for signal intensity and replicate reproducibility. The minimal region of common amplification is represented by a black vertical bar beside the clone positions. (**b**–**d**) FISH validation. The clones used as probes for these experiments are shown in panel **a**, framed in green or red according to their color in the FISH images. (**b**). Tumor 2138: Clones RP11–87N16 and RP11–108B8 were confirmed not to be part of the amplicon (two red and green signals). Blue signals correspond to a probe specific for chr. 8 centromere. (**c**) Tumor 2219: Clone RP11–87N16, with two red signals, was confirmed to be outside of the amplicon and lost in relation to the aneuploid content for chr. 8 (average of five blue signals for chr. 8 centromeric probe). RP11–108B8 was confirmed to be amplified (amplified material shown in green). (**d**) HCC1500 interphase nucleus (left) and metaphase detail (right): Clone RP11–350N15 was confirmed to be excluded from the amplicon (red signals) while CTD-2343B20 amplification was verified (amplified material shown in light blue as a result of the overlapping green of the probe and the dark blue of the chr. 8 paint probe used in this experiment). Metaphase detail shows that the amplified material is part of a rearranged copy of chr. 8

Table 2 Small regions of copy-number loss in chromosome arm 8p

Band	*Start*[b] *(Mb)*	*End*[b] *(Mb)*	*Size*[c] *(Mb)*	*MPL (%)*[a]			*Clones involved*	*Genes in the region*
				Breast tumors	*Breast cell lines*	*Ovarian tumors*		
8p23	2.57	8.58	6.01	28	43	30	RP11-104F14 RP11-45M12 CTD-2629I16 RP11-211C9	 CSM1 DEFB1, DEFA6, DEFA4 CLDN23
8p12	33.39	33.62	0.23	34	33	35	RP11-722E23 RP11-10D7 RP11-2I13	FUT10, LOC84549, FLJ23263 RNF122 MGC1136, LOC442384
	37.99	38.39	0.40	25	10	32	RP11-701H6 RP11-90P5 RP11-350N15	MGC33309, ADRB3, EIF4EBP1 ASH2L, STAR, LSM1, BAG4, DDHD2 WHSC1L1, LOC441345, LETM2, FGFR1

[a]MPL = Mean percentage of loss as described in Materials and methods. The threshold for copy-number loss was set at ratios <0.8 ($\log_2$ ratio <0.32). [b]Midpoints of the clones flanking the region of loss. [c]Calculated as the distance between the midpoints of the clones flanking the region

observed with rtq-PCR, the amplified and nonamplified samples presented a very different pattern of expression in the minimal amplicon and the flanking region. These differences were not found for genes closer to the centromere, which were amplified in only two tumors and were not consistently overexpressed. The genes that presented the highest levels of overexpression were *C8orf2* and *BRF2*, with a median expression increase in the amplified samples of 2.5- and 3.2-fold, respectively. *RAB11FIP1*, *ASH2L* and *LSM1* were also overexpressed, with a median expression ratio in samples with amplification of 2.2 (*RAB11FIP1* and *LSM1)* and 2 (*ASH2L)*. Remarkably, *BRF2*, *C8orf2* and *RAB11FIP1* all map in the minimal region of amplification. *PROSC*, also located within this region, did not show consistent overexpression. *FLJ14299* did not show significant expression changes and *UNC5D* did not show a regular pattern of expression either within the nonamplified or amplified cases. When comparing the expression levels between nonamplified and amplified samples, the differences were significant for *C8orf2*, *BRF2*, *RAB11FIP1*, *ASH2L, STAR*, *LSM1*, *BAG4* and *HTPAP* (Wilcoxon's rank sum test, $P<0.05$). None of the genes located in the centromeric part of the amplicon were significantly overexpressed compared to the nonamplified (Wilcoxon's rank-sum test, $P>0.05$). Importantly, for six out of seven genes analysed using both rtq-PCR and oligo-microarrays, the expression differences between amplified and nonamplified samples were concordant: significant for *RAB11-FIP1*, *LSM1*, *BAG4* and *HTPAP* and nonsignificant for *UNC5D* and *ADAM9*. The seventh gene, *FLJ14299*, was found differentially expressed by rtq-PCR but not by oligo-array analysis.

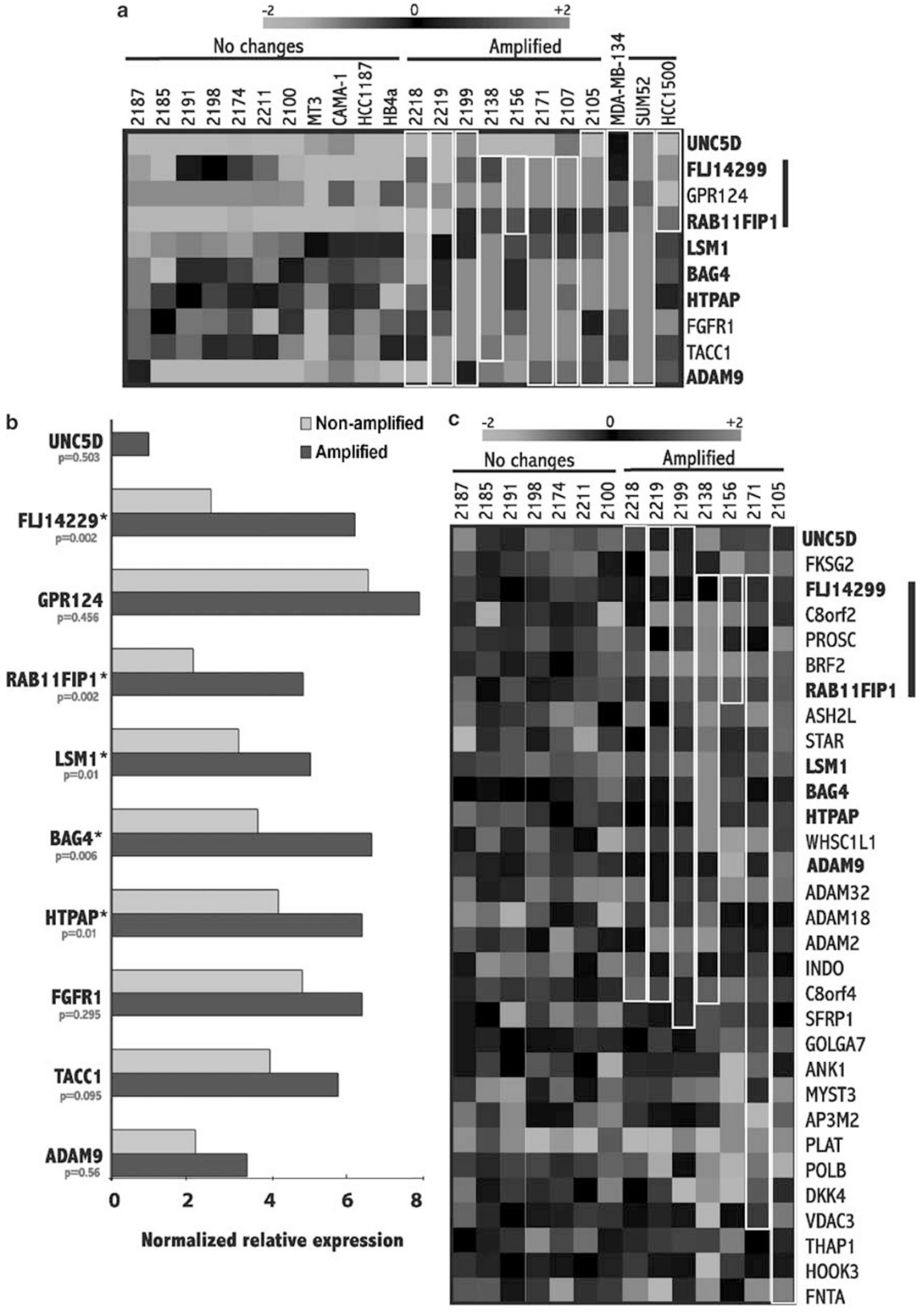
Discussion

The 8p11–12 amplicon has been the subject of numerous studies for many years, but the putative oncogene/s that drive this aberration have yet to be identified. Our long interest in the region and the fact that array-CGH has not been fully used to explore the amplicon prompted us to develop a 8p11–12 high-resolution array and take advantage of the increased coverage to analyse this alteration. The array performance was carefully tested with the appropriate validation experiments. Many cell lines have been extensively characterized by other methods within the group (Courtay-Cahen *et al*., 2000; Davidson *et al*., 2000) (http://www.path.cam.ac.uk/pawefish) and we verified that we could detect previously known changes. Further validation of copy-number changes across several amplified tumors was confirmed by FISH using touchprint preparations.

In our set of breast tumors and breast cell lines, we detected well-known aberrations such as gain of 8q, 1q, 20q or 7p, loss of 17p, 8p, 1p or 18q and amplification of 8p12, 17q12 or 11q13 (Forozan *et al*., 2000). Similarly, in our ovary cases, we observed gain of 8q, 20q, 3q or 1q, loss of 22q, 17p or 9q and amplification at 8q24, changes recurrently observed in these neoplasms (Bayani *et al*., 2002). Therefore, the clones distributed

Figure 4 Gene expression across the 8p11–12 region. (**a**) Expression level of genes located at 8p11–12 assessed by rtq-PCR. Each row corresponds to one gene ordered according to its genomic position from the most distal one (top) to the most proximal one (bottom). Genes analysed by rtq-PCR and oligoarray are shown in bold font. The black bar on the right indicates the genes contained within the minimal region of amplification. Each column represents a sample. Samples are grouped according to whether they show no copy-number changes or amplification at 8p11–12. White frames delimit the genes within the amplicon in each sample. $\log_2$ ratios are depicted in a color scale where red represents upregulation and green downregulation (note that it is the reverse to the CGH display where red conventionally represents loss). (**b**) Comparison of gene expression between samples showing no copy-number changes and amplification at 8p11–12 by rtq-PCR. Bars represent the normalized median $\log_2$ ratio for each group of samples. Asterisks indicate genes whose expression was significantly different between the two groups ($P<0.05$, Wilcoxon's rank-sum test). (**c**) Expression level of genes located at 8p11–12 assessed by expression oligo-microarray (layout as described in **a**)

throughout the whole genome at 10 Mb intervals not only allowed us a proper normalization of the data but also reported patterns of changes in breast and ovarian tumors that reproduced those previously described in literature. These results provided an extra reinforcement of the good performance of the array.



We found amplification of 8p11–12 in 13 cases comprising eight breast primary tumors (24%), four breast cancer cell lines (15%) and one ovary primary tumor (5%). The amplification incidence in our breast tumor series is greater than the 10–15% commonly reported in literature, but remarkably similar to that recently described by Ray *et al*. (2004). These authors found that eight breast tumors out of 32 had high-level gene amplification in part of the 8p11–12 region. Both results taken together suggest that the 8p11–12 amplification might occur in a higher number of breast cancers than published so far. In contrast, 8p11–12 amplification seems to be a rare event in ovarian tumors. We have confirmed the previously reported 8p11–12 amplification in the cell lines SUM52, SUM44 and MDA-MB-134 (Bautista and Theillet, 1998; Ray *et al*., 2004). In addition, we report now that HCC1500 also has 8p11–12 amplification. In fact, its amplicon is not much larger than the minimal region of amplification defined in this study and this cell line should be useful for functional studies of candidate genes.

The size and boundaries of the amplicon were quite heterogeneous, but in most of the samples, the region of amplification was confined within positions 35 and 40 Mb. The *NRG1* gene, therefore, was excluded (except in SUM52), as were the centromere and 8q11. These results show that amplification does not seem to play a major role in the deregulation of *NRG1*, which encodes ligands for the ErbB/EGF-receptor family (including ErbB2/HER2), and we have recently shown it to be the target of breakpoints in 6% of breast cancers (Adelaide *et al*., 2003; Huang *et al*., 2004). Similarly, although it has been suggested that 8q11 might be worthy of further analysis (Ray *et al*., 2004), our results do not support this view.

We were able to define a segment of minimal common amplification of approximately 1 Mb size flanked by clones centered at positions 36.9 and 37.9 Mb. Previously proposed candidates genes such *LSM1*, *BAG4*, *HTPAP*, *FGFR1*, *TACC1* or *SFRP1* (Ugolini *et al*., 1999; Ray *et al*., 2004) map beyond the proximal boundary of the minimal amplicon. The defined core of common amplification contains the genes *FLJ14299*, *C8orf2*, *PROSC*, *GPR124*, *BRF2* and *RAB11FIP1*. *FLJ14299* is a novel gene containing zinc-finger domains that are protein motifs for nucleic acid recognition and these motifs have been found in a number of tumor-related genes such as *WT1* or *BCL6* (Ladomery and Dellaire, 2002). Little is known about *C8orf2*, except for its membrane-associated potential subcellular localization and the fact that is highly conserved in divergent species (Ikegawa *et al*., 1999a). *PROSC* is a highly conserved gene that encodes a protein likely to be in the soluble cytoplasmic fraction, but its function remains to be determined (Ikegawa *et al*., 1999b). *GPR124* encodes the G-protein-coupled receptor 124 or tumor endothelial marker 5 (TEM5) that has been reported to display elevated expression during tumor angiogenesis and neoangiogenesis (Yamamoto *et al*., 2004). *RAB11FIP1* or Rab coupling protein interacts with small GTPases belonging to the Ras superfamily that have regulatory roles in the formation, targeting and fusion of intracellular transport vesicles (Lindsay *et al*., 2002). *BRF2* encodes one of the subunits of the RNA polymerase III transcription factor complex and is specifically required for transcription of the polymerase III-type small nuclear RNA genes (Schramm *et al*., 2000; Cabart and Murphy, 2001). The potential role for snRNAs and dysfunctional snRNAs in the genomic instability of cancer has been considered (Rew, 2003).

When assessing the level of expression of some selected genes distributed along the 8p11–12 region, we found that in most of the amplified cases the genes within the amplicon boundaries were overexpressed. This is in agreement with studies that report the remarkable degree to which variation in gene copy number affects gene expression in breast cancer (Hyman *et al*., 2002; Pollack *et al*., 2002). The concordance between DNA amplification and elevated gene expression highlights the difficulties in determining which gene or genes within an amplicon might play a major role in the oncogenic process. Among the genes localizing in the minimal region of amplification, *FLJ14299* and *GPR124* were the most consistently overexpressed compared to normal luminal cells by rtq-PCR analysis. A recent study showed that *FLJ14299* was amplified and expressed at very high levels in the cell lines SUM44 and SUM225 (Ray *et al*., 2004). We now corroborate this good correlation between amplification and overexpression at this locus in a greater number of samples that include primary tumors and in addition show that *FLJ14299* maps to the newly identified minimal region of amplification. In contrast to the rtq-PCR results, *FLJ14299* was not found overexpressed in the oligo-array experiments. This is likely due to the use of a pool of all the tumors as the common reference for this analysis, which might explain the only slight upregulation in individual tumors relative to the reference. The ability of the rtq-PCR to detect a difference between samples with or without *FLJ14299* amplification is probably due to the wider dynamic range of this method (Yuen *et al*., 2002). *C8orf2* and particularly *BRF2*, both mapping to the minimal region of amplification, showed the highest levels of overexpression in the microarray experiments and showed significant differences in expression when comparing amplified and nonamplified samples. *RAB11FIP1* presented a more moderate level of overexpression, but this was detected by both methods used and found significantly associated with amplification in both analysis. *RAB25* has been recently implicated in the aggressiveness of ovarian and breast carcinomas through amplification (Cheng *et al*., 2004). This suggests that *RAB11FIP1* may also be a candidate. We did not find consistent overexpression of *PROSC*, and *GPR124* high expression levels did not seem to be specifically attributable to amplification.

Outside of the minimal region of amplification, we found that the *UNC5D* gene laid in the vicinity of the amplicon boundary in several samples and that it was the site of a breakpoint in one breast tumor. However, we did not find a pattern of expression consistent with these rearrangements. *FGFR1* was not included in the

minimal region of amplification, and although we found overexpression in multiple samples, this was not specifically associated with amplification. The most centromeric genes in the region were neither consistently found amplified nor overexpressed, which strengthens our findings indicating that the relevant gene/s are located distal to them.

We did not find significant correlation between any of the clinical variables and the presence of 8p11–12 amplification in the breast tumors, which is likely due to the lack of statistical power given the number of analysed cases. More samples will be needed to address this issue and also to elucidate specifically whether the presence of independent amplicons within the region makes a difference when identifying subsets of tumors with different outcome.

In summary, in our study, we have identified a 1 Mb region of common amplification that is likely to contain key genes driving the 8p11–12 amplicon and that does not include other previously suggested candidates. Among the genes within the minimal region of amplification, our data support the exclusion of *GPR124* and *PROSC* as relevant genes. With a good correlation between amplification and overexpression, *FLJ14299*, *C8orf2*, *BRF2* and *RAB11FIP1* are particularly interesting candidates for further functional studies.

Materials and methods

Primary tumors and cell lines

DNA was extracted from frozen samples of 33 breast and 20 ovarian primary tumors and 28 breast cell lines (27 carcinoma derived and one from normal luminal epithelium, detailed in Supplementary Table 1) using standard proteinase K digestion, phenol/chloroform extraction and ethanol precipitation (Sambrook and Russell, 2001). Tumors, all with neoplastic cellularity greater than 50%, were obtained with appropriate ethical approval and are detailed in Supplementary Tables 2 and 3. The breast tumors were primary operable invasive breast cancers from the Nottingham City Hospital Tumor Bank (Elston and Ellis, 1991; Parker *et al*., 2001; Miremadi *et al*., 2002). Ovarian tumors were from Addenbrooke's Hospital, Cambridge.

CGH arrays

Custom arrays were made providing high resolution over the 8p11–12 region (54 clones), average 1.5 Mb coverage for Chr. 8 (82 clones), 34 clones containing 'cancer-related' genes and 10 Mb coverage for the rest of the genome (285 clones). Six *Drosophila* clones from the RPCI-98 library (http://www.chori.org/bacpac/) were included as controls. The 54 BACs covering the 8p11–12 region were spread over 9.5 Mb, between positions 31.03 Mb (RP11–473A17) and 40.58 Mb (RP11–51K12) (http://genome.ucsc.edu). These clones were selected using the Washington University fingerprint map and the UCSC Golden Path draft human genome sequence (http://genome.ucsc.edu). All clones were from the RP11 library, except where indicated and were obtained from the Wellcome Trust Sanger Institute (Hinxton, UK) or from Invitrogen (Paisley, UK). In all, 49 of the 54 clones had been checked for location on 8p by metaphase FISH as described (Courtay-Cahen *et al*., 2000). The clones outside the 8p11–12 region were BACs from the 1 Mb clone set provided by the Wellcome Trust Sanger Institute (Fiegler *et al*., 2003) (http://www.sanger.ac.uk). The BAC DNA was isolated using micropreps and amplified by degenerate oligonucleotide primed PCR (DOP-PCR) as described (Fiegler *et al*., 2003), ethanol precipitated, dissolved in 150 mM phosphate buffer at approximately 300 ng/μl and spotted in triplicate onto amine-binding slides (CodeLink Activated Slides, Amersham Biosciences, Buckinghamshire, UK) using a MicroGrid II arrayer (BioRobotics, Boston, MA, USA).

DNA labeling and hybridization

Labeling and hybridization protocols were followed as described by Fiegler *et al*. (2003) with slight modifications. Volumes were scaled down for an array surface of 2 cm $\times$ 2 cm and prehybridization and hybridization of the slide were performed in an open well created by sticking an adhesive plastic frame delimiting the array area. Evaporation was prevented by placing the slides in humid hybridization chambers (Camlab Ltd, Cambridge, UK). Slides were washed in PBS/0.05% Tween-20 for 10 min at room temperature before and after a main wash in 50%formamide/0.5xSSC for 30 min at 42°C.

Image acquisition, data analysis and array-CGH validation

The arrays were scanned on an Axon 4100A scanner (Axon Instruments, Union City, CA, USA). GenePix Pro 4.1 software (Axon Instruments) was used to perform segmentation and calculate intensities after background subtraction. The output file was imported into an Excel spreadsheet and spots with intensity below twice the median intensity of the *Drosophila* clones were rejected. Test/reference ratios were then calculated and normalized to the median ratio of the autosomal chromosome clones. Spots with ratios more than 10% different from the median of the triplicate were rejected. If a minimum of two spots of the triplicate were accepted, the mean of the $\log_2$ ratios was calculated and plotted against position in the chromosomes according to the NCBI Build 35 of the Human Genome Sequence.

Cell lines that had been previously well characterized by 24-color karyotyping and metaphase FISH (Courtay-Cahen *et al*., 2000; Davidson *et al*., 2000; Adelaide *et al*., 2003) were used to assess the ability of the array to detect copy-number changes. We also included cell lines such as GM04626 and GM0141 with three and four copies of the chromosome X, respectively (Coriell Institute for Medical Research), to validate the array. In male *vs* female, self *vs* self (female *vs* female), GM04626 *vs* female and GM0141 *vs* female hybridizations, the average ratio for the X chromosome clones was 0.64 ± 0.11, 1.02 ± 0.06, 1.33 ± 0.06 and 1.64 ± 0.09, respectively, and the normalized average ratio for the autosomal clones was 1 ± 0.05. This showed the capacity of the array to resolve copy-number alterations as low as single-copy change. When cut-off ratios of >1.2 ($\log_2$ ratio $= 0.26$) for gain and <0.8 ($\log_2$ ratio $= -0.32$) for loss were used, the false-positive rate in these experiments was below 0.2%. For subsequent analysis, we adopted these thresholds to define gain and loss, respectively, and a cutoff ratio of >2 ($\log_2$ ratio >1) to define amplification.

In order to display the results and identify recurrent changes across samples, we used the CGH-analyzer software (Greshock *et al*., 2004). To identify trends in copy-number gain or loss in a given region, we calculated the mean percentage of gain (MPG) or loss (MPL) per clone as described (Douglas *et al*., 2004).

Touchprint and metaphase FISH

Interphase FISH was performed on imprints made from frozen tumor tissues onto APES (3-amino propyl tri-ethoxysilone)-coated slides. Briefly, tumor imprints were fixed in Carnoy's fixative for 15 min at −20°C and left overnight at room temperature. BACs selected for confirmation of array-CGH results were labeled with either biotin or FITC and the chromosome 8 centromeric probe was labeled with digoxigenin. Labeled probes in the presence of excess human Cot1 DNA were denatured at 72°C for 10 min and left to preanneal at 37°C for 30 min. The imprints were denatured in 70% formamide/2xSSC at 72°C for 2 min and dehydrated through an ice-cold ethanol series (70, 90 and 100%). Hybridization was performed at 37°C for 48 h in a humidified chamber. Probes were detected with FITC-labeled goat anti-rabbit and rabbit anti-FITC antibodies (Sigma), streptavidin-CY5 (Amersham, UK) and anti-digoxigenin-rhodamine (Roche, UK). Slides were imaged using a Zeiss Axioplan 2 fluorescent microscope and the SmartCapture VP software (Digital Scientific, UK). Metaphase spreads of the cell line HCC-1500 were prepared using standard procedures and FISH was performed as described (Courtay-Cahen *et al*., 2000).

Real-time quantitative PCR

Total RNA from 26 frozen breast tumors and 27 breast cell lines was extracted by using TRIreagent (Sigma) and purified using RNAeasy kit (Qiagen) according to manufacturer's protocol. RNA from normal human luminal breast cells was kindly provided by Mike O'Hare (Ludwig Institute for Cancer Research/UCL Breast Cancer Laboratory, London, UK). Details about the separation and purification of the luminal breast cells compartment and protocol used for RNA preparation from this cell population can be found elsewhere (Clarke *et al*., 1994; Jones *et al*., 2004). cDNA was synthesized from 2 μg of total RNA (Reverse transcription kit, Applied Biosystems) using random hexamers. Specific primers were designed for amplifying 10 selected genes located in the 8p11–12 amplicon (*UNC5D*, *FLJ14299*, *GPR124*, *RAB11FIP1*, *LSM1*, *BAG4*, *HTPAP*, *FGFR1*, *TACC1*, *ADAM9*) and control genes (*SDHA*, *GAPDH*, *UBC*). rtq-PCRs were performed in triplicate in 10 μl reactions containing 5 μl of SYBR® Green PCR Master Mix (Applied Biosystems), 2 pmol of primers and 1.5 μl of cDNA as template. The cycling conditions for the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) were 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 s, 57°C for 30 s and 72°C for 30 s. Relative expression levels were calculated based on the difference in *Ct* values between the test samples (labeled *s*) and normal breast luminal cells (labeled *c*). This was normalized with expression levels of the control genes (labeled *r*) by using the equation $E_t^{Ct_{ct}-Ct_{st}}/E_r^{Ct_{cr}-Ct_{sr}}$ as described (Pfaffl, 2001). In order to assess whether upregulation of gene expression could be directly attributed to DNA amplification, we compared the relative expression of each gene between the amplified cases and a set of samples that did not show any copy-number change at the region. Specifically, we used the Wilcoxon's rank-sum test (Agresti, 2002) to determine whether the difference in expression was statistically significant.

Expression microarray experiments

In all, 29 out of the 33 primary breast tumors analysed for DNA copy-number changes were also analysed for gene expression using a human oligo-microarray. Total RNA extraction, RNA amplification and indirect labeling were performed as described (Naderi *et al*., 2004). Labeled targets of all tumor samples were pooled as a reference and 1.5 μg of Cy3- or Cy5-labelled cRNA from each sample was hybridized against 1.5 μg of pool reference from the opposite dye. Hybridizations were performed in dye-reversal pairs on oligonucleotide microarrays containing 22 575 features (Agilent® Human 1A 60-mer Oligo Microarray) following the manufacturer's instructions. Scanning, feature extraction and normalization were carried out using Agilent G2565BA scanner and Agilent G2567AA software. The $\log_2$ ratios for 31 genes spotted in the array mapping to 8p11–12 including *UNC5D*, *FKSG2*, *FLJ14299*, *C8orf2*, *PROSC*, *BRF2*, *RAB11-FIP1*, *ASH2L*, *STAR*, *LSM1*, *BAG4* and *HTPAP* were retrieved using Spotfire® DecisionSite 8.0 (Spotfire Inc., Somerville, MA, USA). For these genes, we further normalized the $\log_2$ ratios to their mean expression across samples by subtracting the mean expression value from each of the log-ratio measurements (Segal *et al*., 2004).

Acknowledgements
We thank Celine Courtay-Cahen for characterization of the BACs, Rachel Lyman for preparation of cell lines, Ahmed Ahmed for revision of ovarian tumors staging and grading, Ian Giddings and Jeremy Clark for help printing the CGH-array and David Bowtell for reviewing the manuscript. This work was supported by Cancer Research UK (CRUK), Wellbeing, the Biotechnology and Biological Sciences Research Council (BBSRC) and The Cambridge-MIT Institute (CMI). JDB is a CRUK Senior Clinical Research Fellow.

References

Adelaide J, Chaffanet M, Imbert A, Allione F, Geneix J, Popovici C, van Alewijk D, Trapman J, Zeillinger R, Borresen-Dale AL, Lidereau R, Birnbaum D and Pebusque MJ. (1998). *Genes Chromosomes Cancer*, **22,** 186–199.

Adelaide J, Huang HE, Murati A, Alsop AE, Orsetti B, Mozziconacci MJ, Popovici C, Ginestier C, Letessier A, Basset C, Courtay-Cahen C, Jacquemier J, Theillet C, Birnbaum D, Edwards PA and Chaffanet M. (2003). *Genes Chromosomes Cancer*, **37,** 333–345.

Agresti A. (2002). *Categorical Data Analysis*. John Wiley & Sons Inc.: New Jersey.

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW and Pinkel D. (2000). *Nat. Genet.*, **25,** 144–146.

Bautista S and Theillet C. (1998). *Genes Chromosomes Cancer*, **22,** 268–277.

Bayani J, Brenton JD, Macgregor PF, Beheshti B, Albert M, Nallainathan D, Karaskova J, Rosen B, Murphy J, Laframboise S, Zanke B and Squire JA. (2002). *Cancer Res.*, **62,** 3466–3476.

Bhattacharya N, Chunder N, Basu D, Roy A, Mandal S, Majumder J, Roychowdhury S and Panda CK. (2004). *Exp. Mol. Pathol.*, **76,** 264–271.

Borrow J, Stanton Jr VP, Andresen JM, Becher R, Behm FG, Chaganti RS, Civin CI, Disteche C, Dube I, Frischauf AM, Horsman D, Mitelman F, Volinia S, Watmore AE and Housman DE. (1996). *Nat. Genet.*, **14,** 33–41.

Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R and Isaacs WB. (1996). *Genomics*, **35,** 46–54.

Cabart P and Murphy S. (2001). *J. Biol. Chem.*, **276,** 43056–43064.
Cheng KW, Lahad JP, Kuo WL, Lapuk A, Yamada K, Auersperg N, Liu J, Smith-McCune K, Lu KH, Fishman D, Gray JW and Mills GB. (2004). *Nat. Med.*, **10,** 1251–1256.
Clarke C, Titley J, Davies S and O'Hare MJ. (1994). *Epithelial Cell Biol.*, **3,** 38–46.
Courtay-Cahen C, Morris JS and Edwards PA. (2000). *Genomics*, **66,** 15–25.
Davidson JM, Gorringe KL, Chin SF, Orsetti B, Besret C, Courtay-Cahen C, Roberts I, Theillet C, Caldas C, Edwards PA and Morris JS. (2000). *Br. J. Cancer*, **83,** 1309–1317.
Douglas EJ, Fiegler H, Rowan A, Halford S, Bicknell DC, Bodmer W, Tomlinson IP and Carter NP. (2004). *Cancer Res.*, **64,** 4817–4825.
Elston CW and Ellis IO. (1991). *Histopathology*, **19,** 403–410.
Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M and Nakamura Y. (1992). *Cancer Res.*, **52,** 5368–5372.
Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP and Carter NP. (2003). *Genes Chromosomes Cancer*, **36,** 361–374.
Flanagan JM, Healey S, Young J, Whitehall V, Trott DA, Newbold RF and Chenevix-Trench G. (2004). *Genes Chromosomes Cancer*, **40,** 247–260.
Forozan F, Mahlamaki EH, Monni O, Chen Y, Veldman R, Jiang Y, Gooden GC, Ethier SP, Kallioniemi A and Kallioniemi OP. (2000). *Cancer Res.*, **60,** 4519–4525.
Garnis C, Coe BP, Zhang L, Rosin MP and Lam WL. (2004). *Oncogene*, **23,** 2582–2586.
Greshock J, Naylor TL, Margolin A, Diskin S, Cleaver SH, Futreal PA, deJong PJ, Zhao S, Liebman M and Weber BL. (2004). *Genome Res.*, **14,** 179–187.
Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, Nicholson RI and Griffiths K. (1982). *Br. J. Cancer*, **45,** 361–366.
Huang HE, Chin SF, Ginestier C, Bardou VJ, Adelaide J, Iyer NG, Garcia MJ, Pole JC, Callagy GM, Hewitt SM, Gullick WJ, Jacquemier J, Caldas C, Chaffanet M, Birnbaum D and Edwards PA. (2004). *Cancer Res.*, **64,** 6840–6844.
Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP and Kallioniemi A. (2002). *Cancer Res.*, **62,** 6240–6245.
Ikegawa S, Isomura M, Koshizuka Y and Nakamura Y. (1999a). *Cytogenet. Cell Genet*, **85,** 227–231.
Ikegawa S, Isomura M, Koshizuka Y and Nakamura Y. (1999b). *J. Hum. Genet.*, **44,** 337–342.
Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E, Parry S, Budroni M, Palmieri G, Neville AM, O'Hare MJ and Lakhani SR. (2004). *Cancer Res.*, **64,** 3037–3045.
Ladomery M and Dellaire G. (2002). *Ann. Hum. Genet.*, **66,** 331–342.
Lindsay AJ, Hendrick AG, Cantalupo G, Senic-Matuglia F, Goud B, Bucci C and McCaffrey MW. (2002). *J. Biol. Chem.*, **277,** 12190–12199.
Miremadi A, Pinder SE, Lee AH, Bell JA, Paish EC, Wencyk P, Elston CW, Nicholson RI, Blamey RW, Robertson JF and Ellis IO. (2002). *Histopathology*, **40,** 215–222.
Naderi A, Ahmed AA, Barbosa-Morais NL, Aparicio S, Brenton JD and Caldas C. (2004). *BMC Genom*, **5,** 9.
Parker C, Rampaul RS, Pinder SE, Bell JA, Wencyk PM, Blamey RW, Nicholson RI and Robertson JF. (2001). *Br. J. Cancer*, **85,** 1958–1963.
Pfaffl MW. (2001). *Nucleic Acids Res.*, **29,** e45.
Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW and Albertson DG. (1998). *Nat. Genet.*, **20,** 207–211.
Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL and Brown PO. (2002). *Proc. Natl. Acad. Sci. USA*, **99,** 12963–12968.
Pribill I, Speiser P, Leary J, Leodolter S, Hacker NF, Friedlander ML, Birnbaum D, Zeillinger R and Krainer M. (2001). *Cancer Genet. Cytogenet.*, **129,** 23–29.
Ray ME, Yang ZQ, Albertson D, Kleer CG, Washburn JG, Macoska JA and Ethier SP. (2004). *Cancer Res.*, **64,** 40–47.
Rew DA. (2003). *Eur. J. Surg. Oncol.*, **29,** 764–765.
Sambrook J and Russell DW. (2001). *Molecular cloning. A Laboratory Manual*, Vol. 1. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, pp 6.26–26.27.
Schramm L, Pendergrast PS, Sun Y and Hernandez N. (2000). *Genes Dev.*, **14,** 2650–2663.
Segal E, Friedman N, Koller D and Regev A. (2004). *Nat. Genet.*, **36,** 1090–1098.
Theillet C, Adelaide J, Louason G, Bonnet-Dorion F, Jacquemier J, Adnane J, Longy M, Katsaros D, Sismondi P and Gaudray P. (1993). *Genes Chromosomes Cancer*, **7,** 219–226.
Todd JH, Dowle C, Williams MR, Elston CW, Ellis IO, Hinton CP, Blamey RW and Haybittle JL. (1987). *Br. J. Cancer*, **56,** 489–492.
Toomes C, Jackson A, Maguire K, Wood J, Gollin S, Ishwad C, Paterson I, Prime S, Parkinson K, Bell S, Woods G, Markham A, Oliver R, Woodward R, Sloan P, Dixon M, Read A and Thakker N. (2003). *Genes Chromosomes Cancer*, **37,** 132–140.
Ugolini F, Adelaide J, Charafe-Jauffret E, Nguyen C, Jacquemier J, Jordan B, Birnbaum D and Pebusque MJ. (1999). *Oncogene*, **18,** 1903–1910.
Wang XZ, Jolicoeur EM, Conte N, Chaffanet M, Zhang Y, Mozziconacci MJ, Feiner H, Birnbaum D, Pebusque MJ and Ron D. (1999). *Oncogene*, **18,** 5718–5721.
Yamamoto Y, Irie K, Asada M, Mino A, Mandai K and Takai Y. (2004). *Oncogene*, **23,** 3889–3897.
Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ and Sealfon SC. (2002). *Nucleic Acids Res.*, **30,** e48.

Supplementary Information accompanies the paper on Oncogene website (http://www.nature.com/onc)