Patterns of genomic change in residual disease after neoadjuvant chemotherapy for estrogen receptor-positive and HER2-negative breast cancer

Patterns of genomic change in residual disease after neo-adjuvant chemotherapy for estrogen receptor positive and HER2 negative breast cancer

Aikaterini Chatzipli (1) (*)
Hervé Bonnefoi (2) (*)
Gaetan MacGrogan (2)
Julie Sentis (2)
David Cameron (3)
Coralie Poncet (4)
EORTC 10994/BIG 1-00 Consortium (**)
Richard Iggo (2)

Affiliations
(1) Wellcome Sanger Institute, Cambridge, United Kingdom
(2) Institut Bergonié, INSERM U1218, University of Bordeaux, France
(3) University of Edinburgh, UK
(4) EORTC Data Centre, Brussels, Belgium

(*) These authors contributed equally
(**) A list of authors and their affiliations appears at the end of the paper

Corresponding author
Richard Iggo, ORCID 0000-0002-8822-9111
Richard.Iggo@u-bordeaux.fr

Key words: breast cancer, neoadjuvant chemotherapy, next generation sequencing, clonal evolution

Abstract

Background

Treatment of patients with residual disease after neo-adjuvant chemotherapy for breast cancer is an unmet clinical need. We hypothesised that tumour subclones showing expansion in residual disease after chemotherapy would contain mutations conferring drug resistance.

Methods

We studied estrogen receptor and/or progesterone receptor-positive, HER2-negative tumours from 42 patients in the EORTC 10994/BIG 00-01 trial who failed to achieve pathological complete response. Genes commonly mutated in breast cancer were sequenced in pre and post-treatment samples.

Results

Oncogenic driver mutations were commonest in PIK3CA (38% of tumours), GATA3 (29%), CDH1 (17%), TP53 (17%) and CBFB (12%); and amplification was commonest for CCND1 (26% of tumours) and FGFR1 (26%). The variant allele fraction frequently changed after treatment, indicating that subclones had expanded and contracted, but there were changes in both directions for all of the commonly mutated genes.

Conclusions

We found no evidence that expansion of clones containing recurrent oncogenic driver mutations is responsible for resistance to neoadjuvant chemotherapy. The persistence of classic oncogenic mutations in pathways for which targeted therapies are now available highlights their importance as drug targets in patients who have failed chemotherapy but provides no support for a direct role of driver oncogenes in resistance to chemotherapy.

ClinicalTrials.gov: EORTC 10994/BIG 1-00 Trial registration number NCT00017095.

Background

Neoadjuvant chemotherapy (NAC) before loco-regional treatment is widely used in patients with large operable or locally advanced breast cancers. Several trials have demonstrated that patients without residual invasive disease in the surgical specimen (primary breast tumour and lymph nodes) after neoadjuvant chemotherapy have a low risk of relapse [1-3] and trials aiming to de-escalate therapy in this group of patients are ongoing. On the other hand, patients with residual invasive tumour after neoadjuvant chemotherapy have at least a 25% risk of invasive relapse within 5 years [1-4]. This high risk of relapse is observed in all molecular breast cancer subtypes despite the systematic use of adjuvant hormonal therapy in luminal tumours and trastuzumab in HER2-positive tumours [1-4].

Based on these data, treatment of patients with residual invasive disease after NAC should be considered as an unmet clinical need. Several studies have been conducted or are ongoing to assess the clinical impact of additional post-neoadjuvant treatments in these patients after standard neoadjuvant treatment. This approach has been successful in two recently published phase III studies. A first study conducted in luminal HER2-negative and triple negative breast cancers (TNBC) compared observation to capecitabine for six to eight cycles [5]. This trial demonstrated a survival advantage in the overall population which was more pronounced in the TNBC group. A second study conducted in HER2-positive tumours compared standard treatment with trastuzumab for 1 year to trastuzumab emtansine (T-DM1) for the same duration [6]. This study demonstrated a halving of the risk of invasive relapse with T-DM1 as compared to continued trastuzumab.

Although this "one size fits all" post-neoadjuvant chemotherapy strategy in patients with residual disease has been successful in these two phase III studies, a large proportion of patients will still relapse. A personalised post-neoadjuvant chemotherapy strategy based on genomic analysis of residual disease could have two important advantages. First, tumour cells in the residual tumour after neoadjuvant chemotherapy are by definition resistant to this treatment. We would expect genomic aberrations conferring resistance to be more easily identified in samples taken after treatment had eliminated the drug-sensitive cells present in the primary tumour. Second, genomic analysis of tumour cells in residual disease could identify de novo mutations that were not present before chemotherapy. These mutations might create new therapeutic opportunities. A previous study focusing on residual disease after NAC for TNBC identified genetic alterations which could potentially be targeted in 90% of the tumours [7] but there was no evidence that these were de novo mutations selected by the treatment. Another study on TNBC and one including multiple subtypes also failed to identify any recurrent changes after treatment [8,9].

In this study, we chose to focus on luminal HER2-negative tumours because these tumours are the most frequent and account for the largest number of breast cancer deaths. We therefore believe that the greatest unmet need lies in this group. The genomic landscape of luminal (ER positive and HER2 negative) tumours has been reported in the literature both before any systemic therapy [10] and in the metastatic setting [11]. However, to our knowledge, there are no data available from this group of tumours after NAC. We were particularly interested in the possibility that post-treatment samples would show expansion of clones containing mutations that can be targeted by therapies that have shown efficacy in patients with metastatic disease, such as PIK3CA mutations [12], BRCA1/2 mutations [13] and AKT1 mutations [14].

Formally, the primary objective of this study was to discover mechanisms of resistance to neoadjuvant chemotherapy in luminal breast cancer by identifying genetic changes in subclones showing clonal expansion after treatment. To identify clonal expansion we analysed variant allele fractions (VAF) in matched samples before and after chemotherapy from 42 patients with luminal (ER and/or PgR positive) HER2-negative breast cancer who did not achieve a pathological complete response (pCR) after neoadjuvant therapy within the EORTC 10994/BIG1-00 trial [15].

The neoadjuvant design of the original 10994 study means we can compare pre-treatment with post-treatment samples to identify changes potentially selected by the treatment. This approach reduces confounding by germline mutations. To reduce the background from passenger mutations we restricted the analysis to a panel of genes that are known to be mutated in breast cancer.

Patients and methods

Study design, treatment and eligibility

This is an ancillary study of the EORTC 10994/BIG 1-00 trial which randomized patients between six cycles of a non-taxane regimen (5-fluorouracil, epirubicin, cyclophosphamide) and a taxane-based regimen (docetaxel for three cycles followed by epirubicin + docetaxel for three cycles) all given before primary surgery [15]. Eligible patients were women aged less than 71 years with histologically-proven invasive carcinoma of the breast suitable for neoadjuvant chemotherapy, with any large operable or locally advanced/inflammatory breast cancer. The trial was approved by national and/or local ethics committees in all participating centres. Patients gave signed informed consent at the time of enrolment for the original EORTC 10994 study.

For the sub-study that is the subject of this report, we have selected patients in a subgroup of the initial population of 1856 based on the following eligibility criteria: (1) patients eligible for the main p53 trial; (2) patients who gave informed consent for optional research on tumour samples; (3) luminal tumours (ER and/or PgR positive) and HER2-negative based on assessment by local laboratory; (4) patients who received at least one cycle of neo-adjuvant chemotherapy and who did not receive radiotherapy before surgery; (5) non pCR after neoadjuvant chemotherapy; (6) Pre- and post-chemotherapy samples with more than 50% invasive tumour cells and 250-300 ng DNA available (for details see Sup Table 1 and the CONSORT diagram in Sup Fig 1).

Objectives

The primary objective was to discover mechanisms of resistance to NAC by identifying recurrent genetic changes in subclones showing clonal expansion after treatment. We defined clonal expansion as a significant change by Fisher test in the variant allele fraction (VAF) after treatment. The secondary objectives were: 1. To describe the mutational landscape of all the samples (number of coding mutations, number of driver mutations, type of mutations), 2. To identify changes in the mutational landscape after NAC (VAF changes in driver oncogenes, copy number changes in driver oncogenes), 3. To document the expansion and contraction of subclones after NAC.

Tumour samples and pathology assessment

Tumour biopsies were taken with a 14G trucut needle before starting neoadjuvant chemotherapy. Some of these biopsies were fixed in formalin and embedded in paraffin (FFPE) as part of routine assessment, the remaining biopsies were frozen as part of the mandatory procedure in the original EORTC 10994/BIG 1-00 trial. Grade, ER, PR, and HER2 status were assessed by local pathologists from a biopsy taken at diagnosis before starting neoadjuvant chemotherapy and prospectively recorded in the case report form. ER and PR, assessed by immunohistochemistry (IHC), were reported as positive or negative according to each centre's local definition. HER2

negativity was defined as either absence of HER2 gene amplification by fluorescent in situ hybridization or a score of 0-1+ by IHC. Pathological response after completion of neoadjuvant chemotherapy was assessed by local pathologists on formalin fixed samples taken during surgical resection of the residual tumour. Non pCR cases were defined as patients with residual invasive disease in the primary tumour and/or in the nodes as reported by the local pathologist.

We collected and centralized in Bordeaux both frozen pre-chemotherapy samples and FFPE post-chemotherapy samples taken during surgery. We have summarised in Sup Fig 1 how the 42 patients were selected from the initial population of 1856 patients, of whom 499 with ER and/or PgR positive and HER2-negative tumours were theoretically eligible for this substudy. We initially considered performing NGS on pre-chemotherapy FFPE trucut biopsies but a pilot study showed that NGS was not feasible with these samples. Post-chemotherapy samples were taken from surgical specimens and formalin fixed. H&E sections were examined to estimate tumour cellularity. All of these samples were centrally reviewed by a board-certified pathologist (GMG) to estimate the tumour cell content and to mark regions in the FFPE samples for DNA extraction. The samples qualified if more than 50% of the cells in the biopsy sample were invasive cancer cells.

DNA extraction and sequencing

In Bordeaux (Bergonié Institute), DNA was purified on Qiagen columns after proteinase K digestion of the pre-chemotherapy frozen samples. For post-chemotherapy FFPE samples, one or two 1 mm diameter cores of 0.5 to 1 mm thickness, depending on the material available, were taken and DNA was extracted with the GeneRead DNA FFPE kit (ref 180134 Qiagen, Hilden, Germany). DNA concentration was measured by Qubit and tubes containing 250-300ng of DNA at >30ng/ul were sent to the Wellcome Sanger Institute (WSI). An electronic manifest linking barcodes to EORTC 10994 Sample IDs was sent to WSI with the tubes.

In Cambridge (WSI), samples were further anonymised with replacement of the EORTC Sample ID with a Sanger ID. 100 SNPs were used to check that pre and post-treatment samples were derived from the same patient. Genomic DNA was quality controlled and 200 ng was used for library preparation. DNA was sheared to an average fragment size of 150 bp using the E210 Covaris plate system (Covaris, Inc. Woburn, MA). The fragmentation settings used were 200 cycles per burst at intensity 4 for 120 seconds. Fragmented DNA was subjected to Illumina DNA sequencing library preparation using a Bravo automated liquid handling platform.

Sequencing libraries were amplified using the bridge-amplification process by Illumina HiSeq pair read cluster generation kits (TruSeq PE Cluster Kit v2.5, Illumina) and were hybridized to custom RNA baits for the Agilent SureSelect protocol. Paired-end, 75 bp sequence reads were generated using Illumina HiSeq 2000 with the target of 1Gb sequence per sample. The sequenced reads were aligned to the reference human genome (NCBI build 37) and deposited in bam format

at the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega/, cram files EGAD00001003334, study accession number EGAS00001001223). Additionally, bam files for 93 normal samples sequenced with the two panels used in this study were deposited under the accession numbers EGAS00001002124 and EGAS00001001808 (panels V3 and V4, respectively).

Bioinformatic Analysis

The breast cancer panels used for DNA capture are described in supplementary data files 1 & 2. They target all coding exons of 366 (first batch) and 280 (second batch) cancer genes that are recurrently mutated, amplified or deleted plus regions showing recurrent copy number variation, and recurrently rearranged introns of fusion genes. The coverage of the first panel is 2028799 bp, that of the second is 2055503 bp, and the overlap is 1089744 bp. Sequencing reads were aligned to the NCBI build 37 human genome using the Burrows-Wheeler Aligner (BWA-aln version 0.5.9, [16] to create BAM files with Smith-Waterman correction [http://bio-bwa.sourceforge.net]. PCR duplicates were marked with biobambam version 2.0.65 [https://github.com/gt1/biobambam2] [17].The tumour samples were compared to 93 FFPE-treated normal samples sequenced with the two panels to an equivalent depth. The sample mean sequencing depth was 574x in batch 1 and 787x in batch 2. The mean coverage with 100+ reads was 82% in batch1 and 97% in batch 2. Coverage for individual samples is shown in Sup Fig 2. Variant calling was performed using the Pindel version 4.2 [https://github.com/genome/pindel] and CaVEMan version 1.5.3 [https://github.com/cancerit/CaVEMan] algorithms as described [18]. A simple tandem repeat filter was applied first to remove variants observed less than five times or in less than 10% of the reads. Variants present in the normal samples were excluded if the difference in VAF was <0.2. Variants were only considered if observed in both forward and reverse strands. The calls were limited to the regions specified in the panel bed files (sup data 3 & 4). The probes generate sequence up to 50 bp on either side of each exon in order to detect splice site mutations. Synonymous mutations in coding regions and variants in UTRs, enhancers and copy number regions are not scored as coding variants. Together with biological selection, for example, counterselection of inactivating mutations in oncogenes, this explains why only 10% of variants are coding mutations. The CNVKit algorithm version 0.9.7b1 was used for copy number segmentation [https://github.com/etal/cnvkit]. Unfortunately, the algorithm was unable to fully correct for the lower quality of the FFPE data, leading to batch effects that are visible in the copy number heatmap . To avoid overinterpreting copy number changes we only scored genes as amplified if the affected segment was <15 Mb. For the reported CCND1 amplicons, the median and maximum lengths of the amplified segments were 3.3 and 12.8 Mb; for FGFR1 they were 3.5 and 14.6 Mb. To enrich for high-confidence somatic variants, further filtering was performed to remove known constitutional polymorphisms using human variation databases (Ensembl GRCh37, 1000 genomes release 2.2.2 and ESP6500, cut-off

VAF ≤0.001); and to remove alterations seen recurrently in the 93 normal DNA samples sequenced with the same protocol. The Variant Allele Fraction was corrected for normal tissue contamination with vafCorrect version 5.4.0 [https://github.com/cancerit/vafCorrect]. Enrichments are defined as variants detected in pre-chemotherapy samples and increased in post-chemotherapy samples. Gains are defined as variants not detected in pre-chemotherapy samples but detected in post-chemotherapy samples. To distinguish between de novo mutations and clonal expansion, coding variants giving VAF = 0 in one sample from a patient were called in both samples without filtering: of the 20 cases where the variant was present in the bam files but removed by the filters, 8 were present once, 3 present twice, 2 present 3 times, 1 present 4 times, 2 present 5 times, 2 present 8 times and 2 present 10 times (Sup Table 2). Code availability: Code is available on request to AC or RI.

Results

Demographics, tumour characteristics and treatment of the 42 eligible patients included in this study are given in Sup Table 1. DNA was extracted from frozen biopsies before treatment and from formalin-fixed samples after treatment. It was sequenced in two batches: an initial batch of 17 tumours with single pre-treatment and post-treatment samples; then a second batch including 18 tumours with single pre-treatment and post-treatment samples and 7 tumours with two pre-treatment samples and one post-treatment sample. Thus 91 samples (49 pre-neoadjuvant chemotherapy and 42 post-neoadjuvant chemotherapy) from a total of 42 tumours had evaluable NGS and CNA data.

Mutational landscape

The total number of SNVs and indels in each sample is shown in Sup Fig 3; the median number of variants was 86, with five samples having over 200 variants. As expected, the vast majority of changes in all samples were C>T and G>A substitutions (Sup Fig 4). All three samples from patient PD30315 had a mutation rate much higher than the samples from the other patients. This tumour has an S314A mutation in POLE (DNA polymerase epsilon) that may explain the high mutation rate. It lies in the exonuclease (proofreading) domain and is flanked by known mutations (E311D and D316G, http://mutationaligner.org/domains/PF03104?gene=POLE). It is rare in the EXAC SNP database (8×10^{-5}), and multiple computational predictions score it as pathogenic (DANN, FATHMM, LRT, MutationAssessor and MutationTaster; varsome.com/variant/hg19/rs770403791).

Coding variants

The median number of coding variants (SNVs and indels) per sample was 8, with 2 samples having over 20 coding variants and 12 samples containing less than 5 coding variants (Sup Fig 5). The coding variants in individual genes in each tumour are listed in Sup Table 2. Driver mutations were defined as previously characterised oncogenic mutations [19,20] or mutations that truncate a tumour suppressor protein. The number of tumours with driver mutations in the pre-chemotherapy or post-chemotherapy samples is shown in Fig 1. Seventeen genes had driver mutations in more than one tumour. The commonest genes with driver mutations were PIK3CA (16/42 tumours = 38%), GATA3 (12/42 tumours = 29%), CDH1 (7/42 tumours = 17%), TP53 (7/42 tumours = 17%) and CBFB (5/42 tumours = 12%).

Clonal expansion and contraction after neoadjuvant chemotherapy

The primary objective was to discover mechanisms of resistance to NAC by identifying genetic changes in subclones showing clonal expansion after treatment. To identify subclones showing clonal expansion we compared the VAF in pre- and post-treatment samples (Fig 2a). There were

31 putative de novo mutations that appeared after treatment. This was balanced by loss of 30 mutations through clonal extinction. These changes affected 46 genes, including 28 driver mutations, in 30 patients. By far the commonest change was PIK3CA mutation but there were changes in both directions: 3 tumours lost mutations, 2 tumours gained them and one tumour had one change in each direction. The next commonest changes were in KMT2C (1 loss, 2 gains) and ARID1A (2 losses, one gain). Examination of the bam files revealed that in 75% of these cases, the variant was actually present at a very low depth when it was called as absent (in two thirds of these cases it was present in one, two or three reads). While it is difficult to draw definitive conclusions from such small numbers of reads, it is likely that in many cases where a new mutation is detected we are seeing expansion of a pre-existing clone rather than a de novo mutation.

Fisher tests were used to detect significant changes in VAF that fell short of complete gain or loss. Including the complete gains and losses, 36% of coding variants changed significantly after treatment (Supplementary Table 2, 152/428 variants had $p<0.01$ after Bonferroni correction). Fig 2b&c illustrate different patterns of clonal behaviour. In Fig 2b (tumour PD30309) all points lie close to the diagonal, indicating that there was no clonal selection: the cluster at VAF ~0.15 represents a tumour subclone whose abundance did not change; the cluster at VAF ~0.5 represents either an ancestral tumour clone or heterozygous germline variants that escaped filtering. Consistent with this interpretation, the variants found at VAF ~0.15 are classified as driver mutations in CDH1, PIK3CA, TBX3 and MAP3K1, whereas the variants at VAF ~0.5 are heterozygous missense changes in ACAN, FGFR1, HRNR, PTPN11, RPL5, and WBP1 that are probably either ancestral passenger mutations or rare germline variants. Fig 2c (tumour PD26285) illustrates expansion of a malignant subclone containing CBFB, CDH1 and NF1 driver mutations from a VAF ~0.15 before treatment to a VAF ~0.55 after treatment. The RUNX1 mutation that goes from VAF ~0.15 to ~0.3 is potentially a mutation present initially in the same parental subclone but which failed to expand. Interestingly, RUNX1 is the dimerization partner of CBFB. The CBFB-RUNX1 dimer tethers ER to enhancers; inactivating mutations in either gene could redirect ER to more pernicious targets. Since the subclone containing the CBFB mutation expanded faster it would appear either that the CBFB mutation is better able to inactivate the dimer or perhaps the mutations in CDH1 and NF1 gave the CBFB subclone an added advantage. The remaining variants in this tumour, at VAF ~0.6 before and after treatment, are non-driver mutations in TCF7L2 and TSC1 that could represent passenger mutations or rare germline variants. Similar plots for each individual tumour showing the VAF before and after treatment are presented in Sup Fig 6.

Next, instead of looking at all genes in individual tumours, we examined individual genes in all tumours. We give an overview showing the driver oncogenes most frequently mutated before and after treatment in Fig 3a; strikingly, all lie almost perfectly on the diagonal, indicating that none

of these genes is consistently selected for by neoadjuvant chemotherapy. Our interpretation is that they are unlikely to confer resistance to chemotherapy. We can examine the behaviour of individual genes by plotting the VAF for driver mutations in the gene in all tumours (Fig 3b-f). The points represent individual tumours, with the size of the points scaled according to the probability that the VAF changed after treatment. It is immediately obvious that for every gene there are points located far from the diagonal, meaning the clonal abundance changed. Strikingly, there were changes in both directions, meaning driver mutations in a particular gene became more abundant in some tumours and less abundant in others. That clones containing driver oncogenic mutations should become rarer is surprising, particularly for a gene like PIK3CA, but it is important to note that some of the variation may be explained by incomplete correction for the amount of contaminating normal tissue in the samples.

Copy number variants

Copy number changes are plotted in a heatmap in Fig 4a. Gains are shown in red, losses in blue. There are frequent gains of chr1q and chr16p, indicating that the der(1;16) signature translocation of luminal breast cancer was present in most samples. Large chr8q gains, commonly attributed to MYC, were also present in most samples. The most commonly amplified genes were CCND1 (11/42 tumours = 26%) and FGFR1 (11/42 tumours = 26%) (Fig 4b). ERBB2 amplification on chr17 was not seen because tumours with this amplicon were excluded from the study. No fusion genes were detected. Both at the level of the whole genome (Fig 4a), and when individual amplicons were examined (Fig 4b), there were no consistent changes after treatment. There was also no significant difference between the number of breakpoints before and after treatment in the two treatment arms.

Pathway analysis after treatment

In Fig 5 we categorise the driver mutations present after chemotherapy into different functional pathways. The only ESCAT [21] tier IA target is PIK3CA, which has hotspot mutations in 14 tumours. We score the POLE mutation in tumour PD30315 as tier IC, because the tumour has a high mutation burden making it a candidate for immunotherapy. Three tumours had tier IIA PTEN mutations, two tumours had tier IIB AKT1 T17E mutations, two tumours had tier IIIa MDM2 amplification, and one tumour had a tier IIIA somatic BRCA1 mutation. Together, 48% of tumours (20/42) had at least one of these defects. If one adds FGFR1 and CCND1 amplicons, which do not respond to current targeted therapy (tier X), 74% of tumours (31/42) had potentially interesting targets. The remaining mutations in Fig 5 are tier IVA or not in the ESCAT breast cancer list, although many of the latter are targets for preclinical or clinical drugs in other tumour types.

Discussion

Our working hypothesis was that clones with recurrent mutations in driver oncogenes would expand after neoadjuvant chemotherapy but we have found no evidence to support it. Since there were multiple changes in VAF and copy number of driver oncogenes after treatment but they show no consistent pattern we conclude that there is no driver oncogene that undergoes strong positive selection under chemotherapy. We confirmed the presence of targetable mutations in many driver oncogenes previously reported to be mutant in breast cancer, and showed that they are not eradicated by neo-adjuvant chemotherapy, but we did not identify any recurrent genetic changes in clones surviving chemotherapy. We therefore conclude that classic mammary oncogenes are probably not responsible for resistance to chemotherapy. This echoes the conclusion from the original EORTC 10994 "p53 study" [15], that p53 status can not be used to select patients for chemotherapy with taxanes. The likely explanation for the negative results in both studies is that chemotherapy targets dividing cells whose rapid division is a distant consequence of the oncogenic mutations that created the tumour. Crucially, the targets of chemotherapy are never mutant in tumours. This contrasts with the paradigm of modern targeted therapy. Pharmacologists will not be surprised by our results: they have long known the metabolic pitfalls that dictate the exposure and response of tumour cells to chemotherapeutic drugs.

We have confirmed the presence of targetable mutations in many driver oncogenes previously reported to be mutant in pretreatment samples from luminal ER+ breast cancers [10,22,23]. We have shown here that the same mutations are also present in residual disease after chemotherapy. The persistence of classic oncogenic mutations in functionally important pathways in so many tumours after chemotherapy highlights their importance as drug targets. Targeted therapies are either licensed or in development for most of these pathways. ESMO has developed a Scale for Clinical Actionability of molecular Targets (ESCAT) to classify mutations according to the strength of evidence supporting their use in patients with metastatic breast cancer [21]. Tier of evidence IA refers to genetic alterations for which a licensed treatment is known to be effective. The two main contenders are ERBB2 amplification and PIK3CA mutation. PIK3CA was the commonest driver oncogene mutated both before and after chemotherapy in this study. Tumours with ERBB2 amplification in the pretreatment samples were excluded from this study, and new ERBB2 amplicons were not detected after treatment. Tier IC includes immunotherapy for tumours with a high mutation rate. This is mainly caused by mismatch repair defects but can also result from defective proofreading by the replicative polymerase, DNA polymerase epsilon (POLE). Several tumours had MSH2 mutations of unknown effect. One had a POLE mutation that produced a large increase in the mutation burden, making that tumour an obvious candidate for immunotherapy. Tier II includes gene-drug pairs which have shown activity in breast cancer but for which there is currently no proof that they increase survival in prospective studies. In our data, PTEN loss (IIA) and AKT1 mutation (IIB) fall into this category. Tier III covers gene-drug pairs that show activity in

other tumour types, or family members of tier I genes. Two genes fall in this category in our data (BRCA1 and MDM2). The commonest changes we saw are classified as tier IVA, meaning there is good preclinical data to support their exploration in clinical studies. Tier X includes perplexing defects for which treatment should work but does not. It includes CCND1 and FGFR1, the genes most commonly amplified in our study. It is surely only a matter of time before the failure of tumours with these amplicons to respond to ostensibly good drugs is explained, leading to the development of new therapeutic strategies that improve the survival of affected patients. Many of the other mutations in Fig 5 were not assigned a level of evidence because they were not in the 40 genes ESMO chose to classify, but effective preclinical or clinical drugs exist for many of them. Mutations in others, like RB1 and FAT1, provide useful information because they confer resistance to licensed drugs. There is thus good reason to be optimistic about the utility of sequencing data, and to foresee an increasing dependence on it in the future. Considering the high risk of relapse of patients with ER+ HER2- tumours with residual disease after NAC, a personalised post NAC strategy based on genomic analysis should be considered a clinical research priority.

The gene panels we tested cover all currently known mammary oncogenes. The pace of discovery of new oncogenes and tumour suppressor genes in breast cancer has fallen to essentially zero because of the enormous scale of the genomic studies that have already been completed [10,22,23]. This means we can confidently exclude the possibility that a common oncogenic driver was missing from the present study. A whole genome sequencing approach would have allowed us to detect mutations in genes that are not known breast cancer genes, but, had we done this without a normal sample to exclude germline variants, the enormous number of unknown variants discovered would have rendered identification of true mediators of resistance a near impossible task. In cases where the mutant fraction for a particular variant changed in one direction in one tumour, there were changes in the opposite direction in other tumours, suggesting that the differences were not caused by clonal selection by the treatment, but only PIK3CA, TP53, GATA3, CDH1 and CBFB were mutant often enough, and CCND1 and FGFR1 amplified often enough, to warrant this type of analysis. This is a recurring theme in genomic studies of breast cancer: many genes are infrequently mutated leading to a multiplicity of tiny groups that defy statistical inference.

Many drug resistance studies based on clinical trials have produced interesting but inconclusive results [7,8,15,24,25,26]. What is missing is proof that the observed changes play a causal role in drug resistance. To fill this gap requires sophisticated cell biological studies linked to clinical trials. Recent progress in cell culture and xenografting techniques [27-29] mean it is now possible, at least in principle, to perform these studies but the technical challenges are daunting. Specifically, it means establishing cultures of samples taken before treatment, introducing putative genetic changes mediating resistance with viral or CRISPR technology, and showing that the genetically engineered

cells acquire the predicted phenotype. Given the risk of sampling artefacts, this should be done with samples containing the broadest possible representation of clones, and with techniques that allow tracing of individual cells in complex mixtures, for example by barcoding the starting material and analysing large numbers of single cells at each step [8].

In summary, we have performed a targeted sequencing study comparing driver mutations before and after neoadjuvant chemotherapy of luminal breast cancer and found no evidence that expansion of clones containing recurrent driver mutations is responsible for resistance to neoadjuvant chemotherapy. There may be multiple such mechanisms in play, each contributing occasionally to resistance, but new approaches based on functional analysis of material from clinical trials will be required to prove it. Despite our failure to discover new chemotherapy resistance mechanisms, we identified targetable defects both before and after chemotherapy in most of the tumours examined. Future clinical studies should seek to increase the number of licensed drugs available to target these defects

Additional information

Acknowledgements

We thank the patients, doctors and nurses involved in the EORTC 10994/BIG 1-00 study for their generous participation. We thank Ultan MacDermott and the Sanger Sample Management and Sequencing team for processing the specimens for NGS.

Authorship

HB, RI and DC conceived the study. DC, HB and EORTC clinical investigators provided clinical samples. CP analysed the clinical data. GMG analysed the histology. JS extracted the DNA. AC and RI analysed the NGS data and made the figures. RI, HB and AC wrote the manuscript. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Ethics approval and consent to participate

The EORTC 10994 clinical trial was registered with ClinicalTrials.gov number NCT00017095 and approved by national and/or local ethics committees in all participating centres. Before registration, all patients signed an informed consent for the trial and for mandatory p53 gene assessment on tumour samples. Patients involved in this substudy gave consent for additional biological research on their tumour samples. The study was performed in accordance with the Declaration of Helsinki.

Data availability

Clinical data can be accessed through the EORTC data sharing platform (for details see www.eortc.org/data-sharing). NGS data can be accessed through the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega/, cram files EGAD00001003334, study accession number EGAS00001001223). The NGS data is only accessible under a Managed Access agreement (for details see www.ebi.ac.uk/ega/dacs).

Competing interests

The authors declare no conflict of interest.

Funding information

We thank the Fondation pour la lutte contre le cancer et pour des recherches medico-biologiques, the Site de Recherche Intégrée sur le Cancer – Bordeaux Recherche Intégrée Oncologie Grant INCa-DGOS-Inserm 6046, and the Breast Cancer Working Group EBCC12 grant for financial support. The funding sources had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the study

418 data and the final responsibility for the decision to submit for publication. Trial design, conduct, and
419 analysis were done at the EORTC headquarters, the Institut Bergonié and Wellcome Sanger
420 Institute independently from all funding bodies.
421
422

EORTC 10994/BIG 1-00 Consortium (**)

Sophie Abadie-Lacourtoisie (5)
Hervé Bonnefoi (2)
Alexandre Bodmer (6)
Etienne Brain (7)
David Cameron (3)
Tanja Cufer (8)
Mario Campone (9)
Elisabeth Luporsi (10)
Cristian Moldovan (11)
Thierry Petit (12)
Martine Piccart (13)
Franck Priou (14)
Elsbieta Senkus (15)
Khalil Zaman (16)

Affiliations

(2) Institut Bergonié, INSERM U1218, University of Bordeaux, France
(3) University of Edinburgh, Edinburgh, United Kingdom
(5) Centre Paul Papin Institut de Cancérologie de l'Ouest (ICO), Angers, France
(6) Hôpitaux Universitaires de Genève (HUG), Geneva, Switzerland
(7) Institut Curie-Hôpital René Huguenin, Saint-Cloud, France
(8) Tanja Cufer, Institute of Oncology, Ljubljana, Slovenia
(9) Centre René Gauducheau Institut de Cancérologie de l'Ouest (ICO), Nantes, France
(10) Centre Alexis Vautrin, Nancy, France
(11) Centre Henri Becquerel, Rouen, France
(12) Centre Paul Strauss, Strasbourg, France
(13) Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium
(14) Centre Hospitalier de La Roche/Yon, La Roche/Yon, France
(15) Medical University of Gdansk, Gdansk, Poland
(16) Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland

References

1 Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J. P., Wolmark, N. *et al.* Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* **384**, 164-172 (2014).

2 Bonnefoi, H., Litiere, S., Piccart, M., MacGrogan, G., Fumoleau, P., Brain, E. *et al.* Pathological complete response after neoadjuvant chemotherapy is an independent predictive factor irrespective of simplified breast cancer intrinsic subtypes: a landmark and two-step approach analyses from the EORTC 10994/BIG 1-00 phase III trial. *Ann Oncol* **25**, 1128-1136 (2014).

3 von Minckwitz, G., Untch, M., Blohmer, J. U., Costa, S. D., Eidtmann, H., Fasching, P. A. *et al.* Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol* **30**, 1796-1804 (2012).

4 Bonnefoi, H., MacGrogan, G., Poncet, C., Iggo, R., Bergh, J. & Cameron, D. Molecular apocrine tumours in EORTC 10994/BIG 1-00 phase III study: pathological response after neoadjuvant chemotherapy and clinical outcomes. *Br J Cancer* **120**, 913-921 (2019).

5 Masuda, N., Lee, S. J., Ohtani, S., Im, Y. H., Lee, E. S., Yokota, I. *et al.* Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N Engl J Med* **376**, 2147-2159 (2017).

6 von Minckwitz, G., Huang, C. S., Mano, M. S., Loibl, S., Mamounas, E. P., Untch, M. *et al.* Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. *N Engl J Med* **380**, 617-628 (2019).

7 Balko, J. M., Giltnane, J. M., Wang, K., Schwarz, L. J., Young, C. D., Cook, R. S. *et al.* Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. *Cancer Discov* **4**, 232-245 (2014).

8 Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893 e813 (2018).

9 Yates, L. R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751-759 (2015).

10 TCGA. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).

11 Bertucci, F., Ng, C. K. Y., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N. *et al.* Genomic characterization of metastatic breast cancers. *Nature* **569**, 560-564 (2019).

12 Andre, F., Ciruelos, E., Rubovszky, G., Campone, M., Loibl, S., Rugo, H. S. *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med* **380**, 1929-1940 (2019).

13 Tung, N. M., Robson, M. E., Ventz, S., Santa-Maria, C. A., Nanda, R., Marcom, P. K. *et al.* TBCRC 048: Phase II Study of Olaparib for Metastatic Breast Cancer and Mutations in Homologous Recombination-Related Genes. *J Clin Oncol* 10.1200/JCO.20.02151, JCO2002151 (2020).

14 Hyman, D. M., Smyth, L. M., Donoghue, M. T. A., Westin, S. N., Bedard, P. L., Dean, E. J. *et al.* AKT Inhibition in Solid Tumors With AKT1 Mutations. *J Clin Oncol* **35**, 2251-2259 (2017).

15 Bonnefoi, H., Piccart, M., Bogaerts, J., Mauriac, L., Fumoleau, P., Brain, E. *et al.* TP53 status for prediction of sensitivity to taxane versus non-taxane neoadjuvant chemotherapy in

breast cancer (EORTC 10994/BIG 1-00): a randomised phase 3 trial. *Lancet Oncol* **12**, 527-539 (2011).
16 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
17 Tischler, G. & Leonard, S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine* **9**, 13 (2014).
18 Raine, K. M., Hinton, J., Butler, A. P., Teague, J. W., Davies, H., Tarpey, P. *et al.* cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *Curr Protoc Bioinformatics* **52**, 15 17 11-12 (2015).
19 Chang, M. T., Asthana, S., Gao, S. P., Lee, B. H., Chapman, J. S., Kandoth, C. *et al.* Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* **34**, 155-163 (2016).
20 Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713 (2017).
21 Condorelli, R., Mosele, F., Verret, B., Bachelot, T., Bedard, P. L., Cortes, J. *et al.* Genomic alterations in breast cancer: level of evidence for actionability according to ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). *Ann Oncol* **30**, 365-373 (2019).
22 Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352 (2012).
23 Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
24 Farmer, P., Bonnefoi, H., Anderle, P., Cameron, D., Wirapati, P., Becette, V. *et al.* A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med* **15**, 68-74 (2009).
25 Quenel-Tueux, N., Debled, M., Rudewicz, J., MacGrogan, G., Pulido, M., Mauriac, L. *et al.* Clinical and genomic analysis of a randomised phase II study evaluating anastrozole and fulvestrant in postmenopausal patients treated for large operable or locally advanced hormone-receptor-positive breast cancer. *Br J Cancer* **113**, 585-594 (2015).
26 Weigelt, B., Pusztai, L., Ashworth, A. & Reis-Filho, J. S. Challenges translating breast cancer gene signatures into the clinic. *Nat Rev Clin Oncol* **9**, 58-64 (2011).
27 Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F. *et al.* A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. *Cell* **172**, 373-386 (2018).
28 Sflomos, G., Dormoy, V., Metsalu, T., Jeitziner, R., Battista, L., Scabia, V. *et al.* A Preclinical Model for ERalpha-Positive Breast Cancer Points to the Epithelial Microenvironment as Determinant of Luminal Phenotype and Hormone Response. *Cancer Cell* **29**, 407-422 (2016).
29 Richard, E., Grellety, T., Velasco, V., MacGrogan, G., Bonnefoi, H. & Iggo, R. The mammary ducts create a favourable microenvironment for xenografting of luminal and molecular apocrine breast tumours. *J Pathol* **240**, 256-261 (2016).

Figure legends

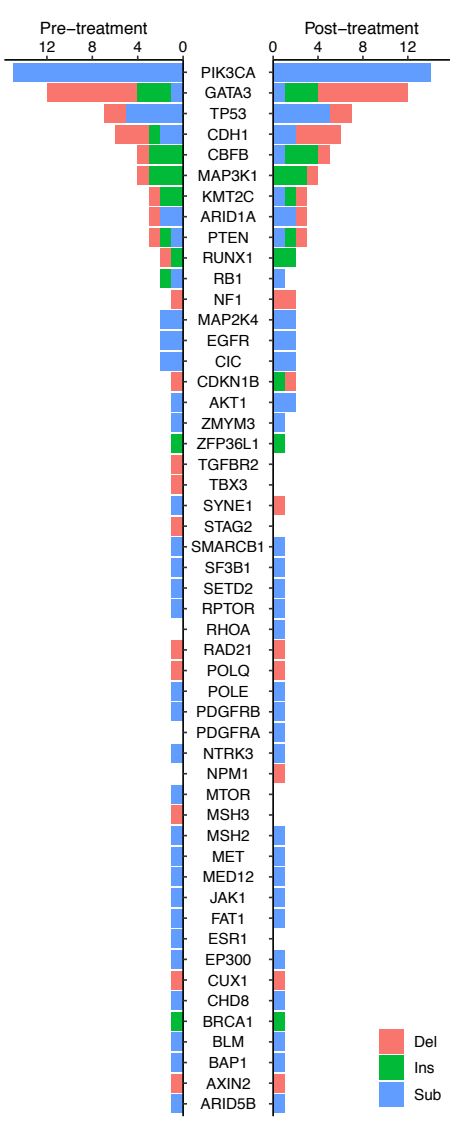
Fig 1. Commonest driver mutations before and after treatment. The mutation count is non-redundant (ie, genes are only counted once if multiple mutations were identified in a tumour, or the same mutation was present in multiple samples from a tumour).

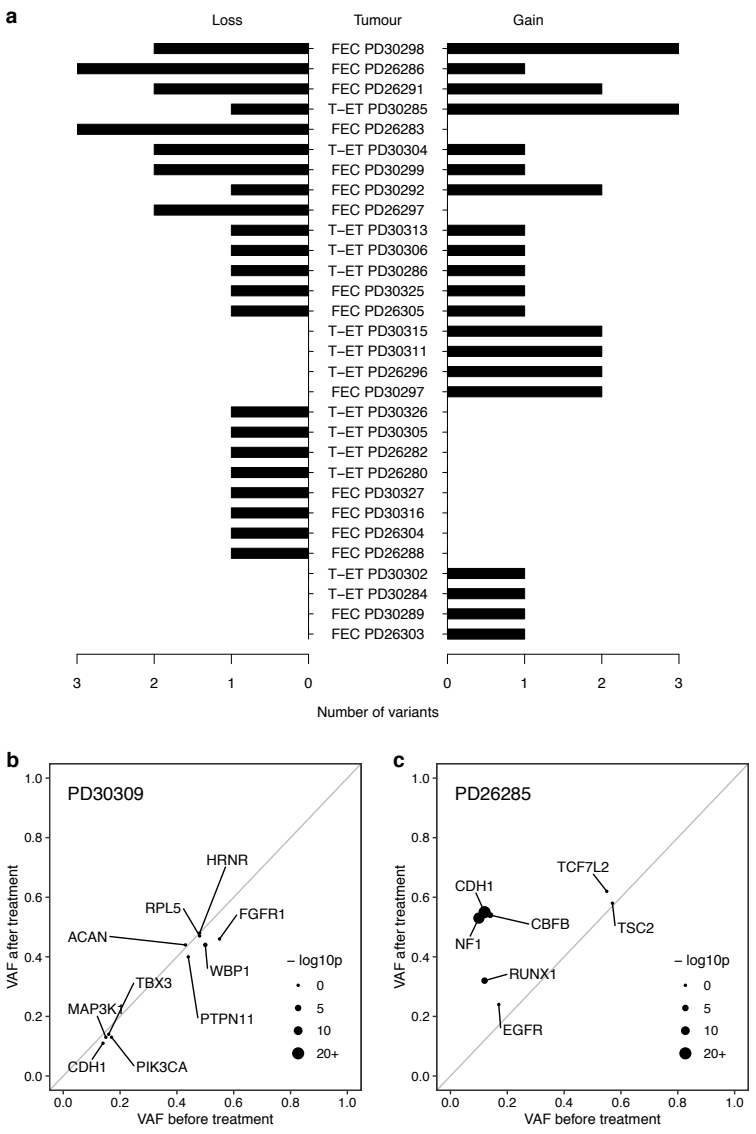
Fig 2. Variants gained and lost after neoadjuvant chemotherapy. a, Number of coding variants gained and lost in each tumour. b&c, Examples of tumours showing different patterns of clonal change after treatment (b, PD30309, FEC arm; c, PD26285, T-ET arm). The size of the plotting symbol reflects the Fisher p value (mutations supported by fewer reads have a smaller symbol).

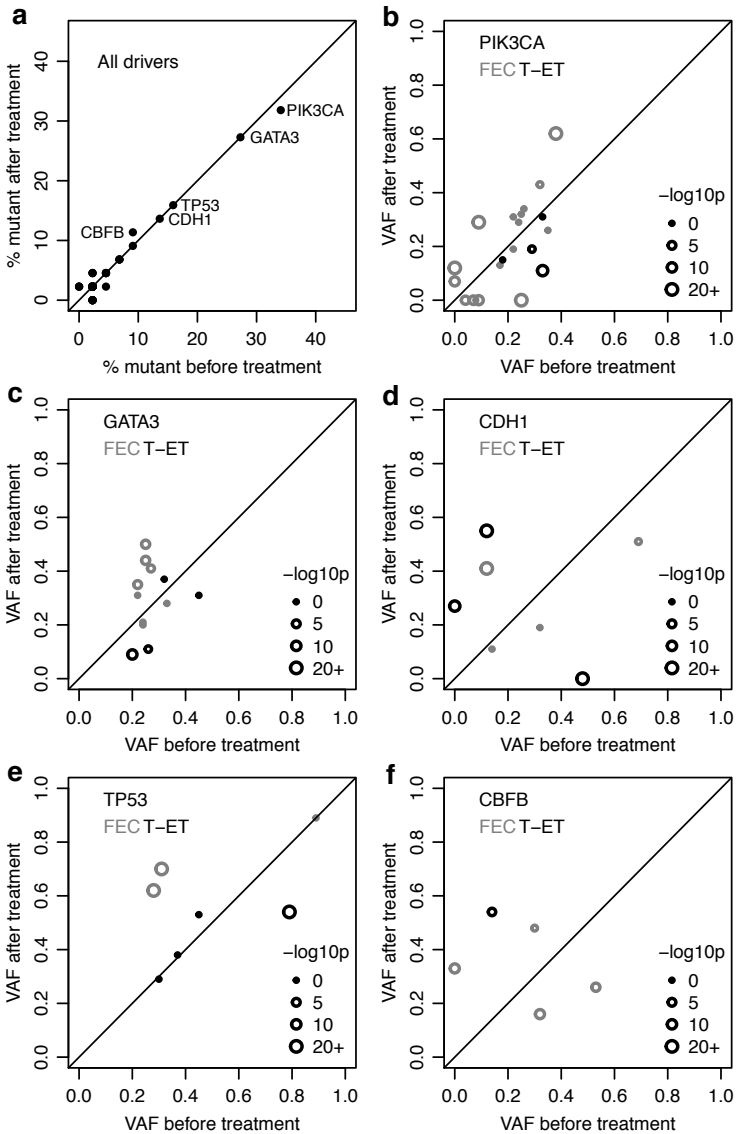
Fig 3. Frequency and VAF of the most frequently mutated oncogenes and tumour suppressor genes before and after neoadjuvant chemotherapy. a, Percentage of tumours with specific genes mutated before and after treatment. b-f, VAF for the most commonly mutated drivers. There are substantial changes in both directions after treatment. The size of the plotting symbol reflects the Fisher p value (mutations supported by fewer reads have a smaller symbol).

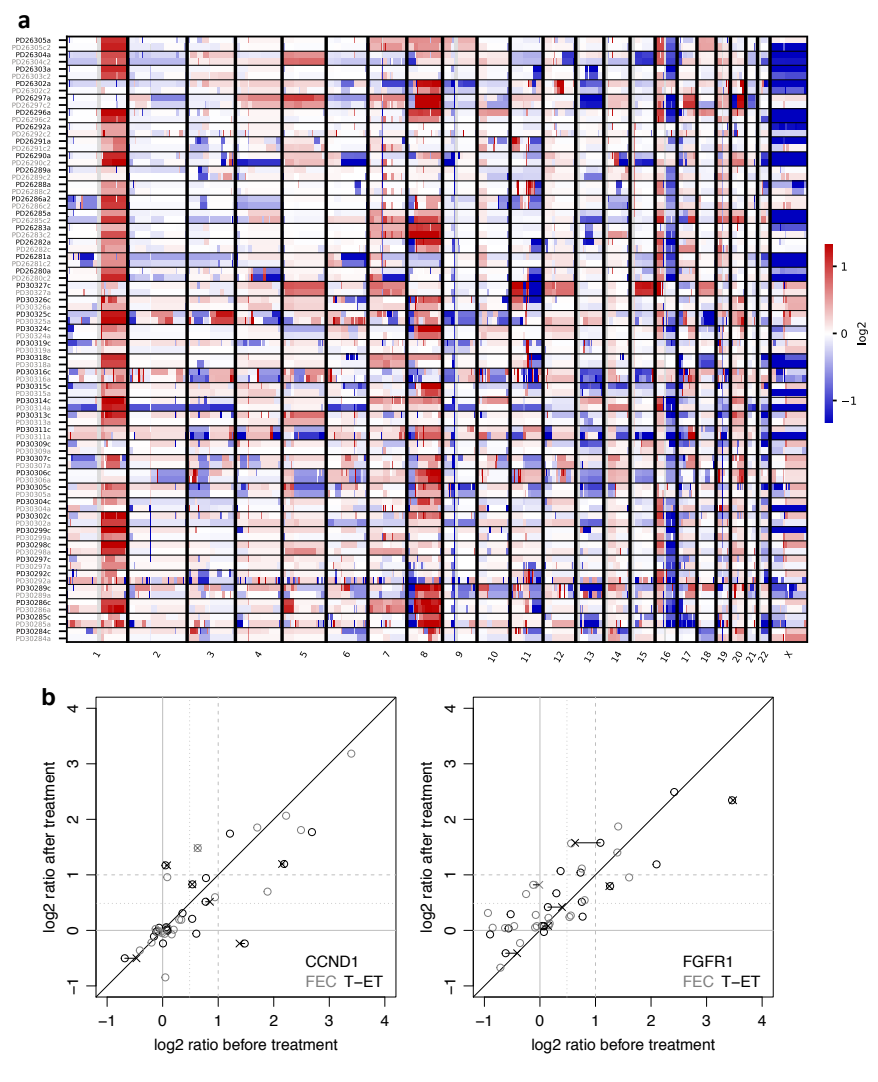
Fig 4. Copy number variants in pre- and post-neoadjuvant chemotherapy samples. a, Heatmap showing regions gained and lost coloured red and blue, respectively. The y axis is ordered by patient ID with pre- followed by post-treatment samples. b, Log2 ratio before and after treatment for the most commonly amplified genes (CCND1 and FGFR1). The dashed grey line corresponds to a copy number of 4. The dotted grey line corresponds to likely gains (3 sd above a copy number of 2). If two pretreatment samples were tested, the second sample is indicated by an X and joined to the first sample by a horizontal line. Note that the FGFR1 sample with log ratio 0.6 pre/1.6 post was not scored as amplified because it was not a focal gain. FEC: 5-fluorouracil + epirubicin + cyclophosphamide x6; T-ET: docetaxel x3 then docetaxel + epirubicin x3.

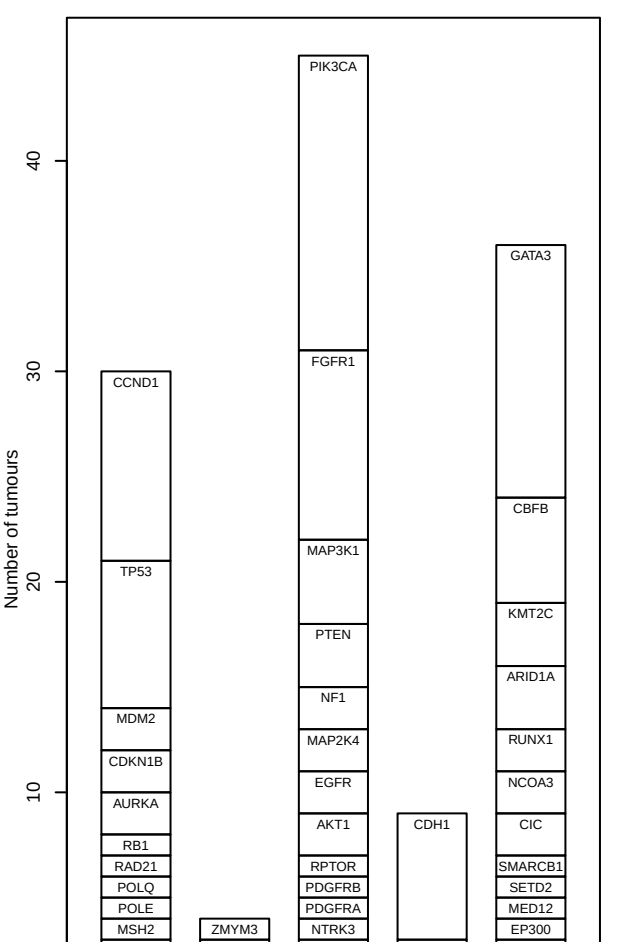
Fig 5. Pathway analysis. Driver oncogenes mutated or amplified after treatment are grouped according to biological pathway. CCND1, FGFR1, AURKA, MDM2 and NCOA3 are within amplified regions; all of the other genes shown contained driver mutations. Note that for amplified regions the true driver may be a nearby gene rather than the one shown.


Pre−treatment
12
8
4
0
Post−treatment
0
4
8
12
PIK3CA
GATA3
TP53
CDH1
CBFB
MAP3K1
KMT2C
ARID1A
PTEN
RUNX1
RB1
NF1
MAP2K4
EGFR
CIC
CDKN1B
AKT1
ZMYM3
ZFP36L1
TGFBR2
TBX3
SYNE1
STAG2
SMARCB1
SF3B1
SETD2
RPTOR
RHOA
RAD21
POLQ
POLE
PDGFRB
PDGFRA
NTRK3
NPM1
MTOR
MSH3
MSH2
MET
MED12
JAK1
FAT1
ESR1
EP300
CUX1
CHD8
BRCA1
BLM
BAP1
AXIN2
ARID5B
Del
Ins
Sub


a
Loss
Tumour
Gain
FEC PD30298
FEC PD26286
FEC PD26291
T–ET PD30285
FEC PD26283
T–ET PD30304
FEC PD30299
FEC PD30292
FEC PD26297
T–ET PD30313
T–ET PD30306
T–ET PD30286
FEC PD30325
FEC PD26305
T–ET PD30315
T–ET PD30311
T–ET PD26296
FEC PD30297
T–ET PD30326
T–ET PD30305
T–ET PD26282
T–ET PD26280
FEC PD30327
FEC PD30316
FEC PD26304
FEC PD26288
T–ET PD30302
T–ET PD30284
FEC PD30289
FEC PD26303
3
2
1
0
0
1
2
3
Number of variants
b
PD30309
HRNR
RPL5
FGFR1
ACAN
WBP1
TBX3
MAP3K1
PTPN11
CDH1
PIK3CA
– log10p
0
5
10
20+
VAF after treatment
VAF before treatment
0.0
0.2
0.4
0.6
0.8
1.0
c
PD26285
TCF7L2
CDH1
CBFB
TSC2
NF1
RUNX1
EGFR
– log10p
0
5
10
20+
VAF after treatment
VAF before treatment


a
All drivers
PIK3CA
GATA3
TP53
CDH1
CBFB
% mutant after treatment
% mutant before treatment
b
PIK3CA
FEC T–ET
VAF after treatment
VAF before treatment
–log10p
0
5
10
20+
c
GATA3
FEC T–ET
d
CDH1
FEC T–ET
e
TP53
FEC T–ET
f
CBFB
FEC T–ET


a
log2
1
0
−1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
b
log2 ratio after treatment
log2 ratio before treatment
CCND1
FEC T−ET
FGFR1
FEC T−ET


Number of tumours
0
10
20
30
40
BAP1
BLM
BRCA1
MSH2
POLE
POLQ
RAD21
RB1
AURKA
CDKN1B
MDM2
TP53
CCND1
NPM1
SF3B1
ZFP36L1
ZMYM3
AXIN2
JAK1
MET
NTRK3
PDGFRA
PDGFRB
RPTOR
AKT1
EGFR
MAP2K4
NF1
PTEN
MAP3K1
FGFR1
PIK3CA
FAT1
RHOA
SYNE1
CDH1
ARID5B
CHD8
CUX1
EP300
MED12
SETD2
SMARCB1
CIC
NCOA3
RUNX1
ARID1A
KMT2C
CBFB
GATA3
Cell cycle
DNA repair
Ribosomes
Splicing
Kinases
Signalling
Cyto−
skeleton
Transcription
Chromatin