Impact of mutational profiles on response of primary oestrogen receptor-positive breast cancers to oestrogen deprivation

Pascal Gellert[1]†, Corrinne V Segal[1]†, Qiong Gao[1], Elena López-Knowles[1], Lesley-Ann Martin[1], Andrew Dodson[1], Tiandao Li[2], Christopher A Miller[2], Charles Lu[2], Elaine R Mardis[2], Alexa Gillman[3], James Morden[3], Manuela Graf[1], Kally Sidhu[4], Abigail Evans[5], Michael Shere[6], Christopher Holcombe[7], Stuart A McIntosh[8], Nigel Bundred[9], Anthony Skene[10], William Maxwell[11], John Robertson[12], Judith M Bliss[3], Ian Smith[4], Mitch Dowsett[1,4]* POETIC Trial Management Group and Trialists

Affiliations:

[1]Breast Cancer Now Research Centre at The Institute of Cancer Research, 123 Old Brompton Road, London, SW7 3RP, UK.

[2]McDonnell Genome Institute, Washington University School of Medicine, 4444 Forest Park Blvd., St Louis, 63108, MO, USA.

[3]Clinical Trials and Statistics Unit at The Institute of Cancer Research, 123 Old Brompton Road, London, SW7 3RP, UK.

[4]Royal Marsden Hospital, Fulham Road, London, SW3 6JJ, UK.

[5]Poole General Hospital, Longfleet Road, Dorset, BH15 2JB, UK.

[6]Southmead Hospital, Westbury-on-Trym, Bristol, BS10 5NB, UK.

[7]Royal Liverpool University Hospital, 200 London Road, Liverpool, L3 9TA, UK.

[8]Queen's University Belfast, University Road, Belfast, BT7 1NN, UK.

[9]University Hospital of South Manchester, Education and Research Centre, Southmoor Road, Manchester, M23 9LT, UK.

[10]Royal Bournemouth Hospital, Castle Ln E, Bournemouth, BH7 7DW, UK.

[11]Withybush General Hospital, Fishguard Rd, Haverfordwest, Pembrokeshire, SA61 2PZ, Wales UK.

[12]University of Nottingham, Derby Rd, Nottingham, NG7 2UH, UK.

†These authors contributed equally to this work

*Corresponding author

Abstract:

Presurgical studies allow study of the relationship between mutations and response of estrogen receptor positive (ER+) breast cancer to aromatase inhibitors (AIs) but have been limited to small biopsies. Here in Phase I of this study, we perform exome sequencing on baseline, surgical core-cuts and blood from 60 patients (40 AI treated, 20 Controls). In poor responders (based on Ki67 change) we find significantly more somatic mutations than good responders. Subclones exclusive to baseline or surgical cores occur in approximately 30% of tumours. In Phase II we combine targeted sequencing on another 28 treated patients with Phase I. We find six genes frequently mutated: *PIK3CA*, *TP53*, *CDH1*, *MLL3*, *ABCA13* and *FLG* with 71% concordance between paired cores. *TP53* mutations are associated with poor response. We conclude that multiple biopsies are essential for confident mutational profiling of ER+ breast cancer and *TP53* mutations are associated with resistance to oestrogen deprivation therapy.

Introduction

Assessment of somatic mutations is becoming increasingly important for the management of cancer patients but molecular heterogeneity occurs across many tumors[1]. This variability is of particular interest in relation to the clonal evolution of individual malignancies but it also poses a severe analytical challenge in terms of the degree to which the whole tumour mutational repertoire is represented by limited biopsy material.

In breast cancer there is major interest in the use of pre-surgical studies for assessing the biological effect of therapeutic agents[2], including the impact that the agents may have on the responsiveness of sub-populations and the emergence of subclones resistant to therapy. However, such studies inevitably depend on analyses of sequential, usually core-cut biopsies that often sample <1% of the tumour mass and may therefore provide limited representation of the tumour genotype.

Breast cancer is the most common malignancy in females in western countries and oestrogen receptor positive (ER+) tumours contribute about 60-75% of the disease[3]. Aromatase inhibitors (AIs) are the most effective agents in post-menopausal woman reducing recurrence rates in primary breast cancer patients by c.50%[4]. These agents inhibit aromatase throughout the body by >97% and suppress plasma oestrogen levels to undetectable levels[5]. However, these therapies are not effective in every patient. Hence, identifying the role that mutations play in *de novo* resistance to AIs is an important clinical research goal.

One large pre-surgical study, PeriOperative Endocrine Therapy - Individualising Care (POETIC) trial, randomized 4,486 patients to receive two weeks' non-steroidal AI or no treatment prior to surgery[2]. Biopsies were collected at diagnosis and at surgery to correlate

molecular alterations in the tumours with their antiproliferative response to an AI. This provides the opportunity to identify DNA alterations that are of biological interest in relation to oestrogen responsiveness and of potential clinical importance in relation to AI use[6]. Like other pre-surgical studies, POETIC is potentially affected by within-tumour heterogeneity. The control group of POETIC (no pre-surgical treatment) allows us to study discrepancies between repeat biopsies from the same patients and to evaluate the molecular heterogeneity within the tumours.

In Phase I of the current study we conduct whole exome analysis followed by capture-probe validation of baseline and surgical core-cut biopsies and of whole blood DNA. We select patients from the control group and treated patients at the extreme ends of the Ki67 response spectrum to study. On the exome-wide mutational profile we find a significantly higher mutational load in poor responding patients indicative for multiple resistance mechanism. Over 2 weeks' of treatment we only find minor effects on the mutational profile in terms of mutational load and variant allele fractions. In about 30% of the tumours we detect intra-tumoural heterogeneity with subclones exclusively to one of the core-cut. In Phase II we perform capture-probe sequencing of baseline and surgical core-cut biopsies and whole blood DNA on additional patients. We concentrate our analysis on mutations in 77 breast cancer genes, for which the entire coding-sequence was added to the capture panel. Through integrating the data from Phase I and II we find a reduced suppression of Ki67 within the poor responder group for *TP53*-mutated tumours and therefor a potential marker for poor response to oestrogen deprivation therapy. We show concordant detection of the mutation status of frequently mutated genes in 76% of the cases. Together with the subclonal analysis we conclude that limited tumour material from

core-cuts complicate mutational profiling of ER+ breast cancer. Multiple biopsies are required for confident mutation calling, especially for heterogeneous tumours.

Results

Clinical Cohort

When Phase I was initiated 148 patients from POETIC (CRUK/07/015) had paired baseline and surgical (2 week) RNA*later* preserved samples available. To focus on a comparison between particularly poor responders and good responders, we excluded treated patients with Ki67 decrease between 60 and 75% (n=34, Methods). After quality assessments, we found 60 eligible sample pairs. Our goal was to choose equal numbers of good and poor responders, but in these pairs only 15 poor responders were found. Therefore, all 25 available good responders were included for a set of 40 treated patients. Together with the 20 pairs from the POETIC untreated control group these constituted the 60 patient cohort of Phase I (Fig. 1a and Supplementary Fig. 1). The patient demographics of samples from Phase I are described in Supplementary Table 1.

To increase the statistical power to examine common events in AI-treated patients Phase II was subsequently conducted including sample pairs that had become available during continual conduct of the POETIC trial. From 108 available pairs of RNA*later* preserved samples, we excluded Controls (n=19) and in keeping with Phase I we excluded samples not falling into either the good or poor responder category (n=19). All 18 available poor responding patients were retained even if one sample of the pair did not meet our criteria (12 pairs, 6 singles) together with 10 good responders paired samples selected based on when they were received first in chronological order (Fig. 1b).

The demographics of all 86 patients in this study are described in Table 1.

Mutation discovery in phase I of the study

Whole exome sequencing (WES) was performed on tissues at baseline and at surgery and on blood from 60 patients (180 samples in total) for initial mutation discovery. This achieved a median coverage of 38x (germline 39x, tumour 37x; Supplementary Data 1); 11 tumour samples including both from one patient (P033) were excluded due to low coverage. We identified a total of 6,910 somatic mutations in the remaining tumour samples from 59 patients.

Mutation validation in phase I of the study

To validate the mutations from WES we performed targeted re-sequencing at higher depth on all 59 patients (excluding 11 tumour samples and one blood from patient P033, 168 samples in total) from above (Supplementary Fig. 2). Therefore, we designed a capture-probe panel covering all potential somatic mutations discovered from WES. Additionally, the entire coding region of 77 previously described breast cancer related genes were added to the panel (Supplementary Table 2). Seven samples attained low coverage, however 6 were sequenced successfully a second time (with samples from Phase II, P003 surgery had to be excluded, mean coverage of 7x). The remaining 167 samples had median coverage 105x (germline 110x, tumour 100x; Supplementary Data 1). Of these, 52 were baseline and 56 were surgical samples consisting of 49 pairs: 17 Control, 11 poor and 21 good responder pairs (Table 2).

The targeted re-sequencing validated 4,232 somatic mutations across the 59 patients that were classified as tier 1 (variants in the coding regions of annotated exons, canonical splice sites, and RNA genes). Without counting identical mutations in paired

142 samples the number of validated mutations was 6,283 mutations across 108 tumour
143 samples (Supplementary Fig. 3 and Supplementary Data 2). These affected 3,388 genes; the
144 majority of mutations were missense (63%) or silent (23%) (Fig. 1c). The mean number of
145 mutations per patient with paired exome-sequencing was 79.5 (median 49, interquartile
146 range, 33.0 to 91.5, Fig. 1d).

147 Two patients were outliers based on their low mutation count (≤8 mutations in both
148 baseline and surgical samples) in the target area. There were two other pairs of samples
149 with extreme differences in their mutation counts between baseline and surgery: 1 *vs* 407
150 (P035, Control) and 86 *vs* 596 (P045, good responder). To exclude sequencing bias, these
151 samples were sequenced a second time to over 200x median combined coverage per
152 sample. The plot of variant allele fractions (VAFs) between the two runs showed high
153 correlations (r=0.85-0.92, Pearson correlation) between the replicates indicating high
154 reproducibility (Supplementary Fig. 4c-f). Despite the higher coverage, many mutations
155 were found in only one or other sample of these pairs (Supplementary Fig. 4a-b), suggesting
156 that these discordances may have been due to normal tissue contamination. This is
157 supported by tumour purity estimation on WES data of these samples (Supplementary Fig.
158 4).

159

160 *Mutational load from phase I samples*

161 For samples in Phase I all potential somatic mutations discovered by WES were
162 added to the capture-panel for validation. This allowed an evaluation of their exome-wide
163 mutational load (that is their total number of mutations). At baseline and at surgery, there
164 was a significant higher mutational load in samples from poor than good responders

(median 62.0 *vs* 33.5, p=0.047, Fig. 2a and median 56.5 *vs* 29.0, p=0.022, Fig. 2b, Mann-Whitney test). Controls showed similar mutation numbers to good responders. There was no significant difference between baseline and surgical samples in mutation counts within the good responders, poor responders or Control (Fig. 2d). However, considering all 32 treated pairs as a group there was a minor but statistically lower median count after treatment (median baseline 43.5 *vs.* surgery 37.0, median of differences -2, p=0.019, Wilcoxon signed-rank test). This significance was retained but weaker after exclusion of the two patients with extreme differences (P035 and P045) from the analysis (p=0.034). Given that the treatment-related differences between baseline and surgery were minor, we merged the mutations in each of the pairs of samples and created a count of unique mutations per tumour giving a value for 49 tumours. Similar to the comparisons described above and shown in Fig. 2a and b, we found that poor responders had significantly more mutations than good responders (median 104 *vs* 41, p=0.021, Fig. 2c, Mann-Whitney test).

We compared the VAFs of mutations between the baseline and surgical sample in all tumours and observed correlations up to 0.86 (Pearson correlation, Supplementary Fig. 5). The VAFs of mutations found in both samples of a pair were significantly lower in surgical than baseline samples for good (median baseline 29.2 *vs* surgery 26.3, p<0.001, Wilcoxon signed-rank test) and poor responders (27.0 *vs* 24.7, p<0.001) but not Control pairs (27.0 *vs* 26.5, p=0.573, Fig. 2e).

Mutational clusters from phase I of the study

We compared the VAFs between baseline and surgical samples to identify mutational clusters which may represent subclones using SciClone[7] (Methods). SciClone

188 analysis was possible in 40 cases: 11 Controls, 20 good and 9 poor responders
189 (Supplementary Fig. 6-8). The median number of identified clusters was 3; the maximum
190 number was 6. Five examples are shown in Fig. 3 selected based on a relatively large
191 number of clusters. We did not perform statistical comparisons of the cluster number
192 between the responder groups because of the small sample size. Visual inspection and
193 comparison of SciClone plots did not reveal differences in the degree of heterogeneity
194 between good and poor responders with both groups having patients showing low and high
195 heterogenic sample pairs. In most pairs, the clusters were represented in both samples of
196 the pair (e.g. P007, P014 and P039, Fig. 3). In about 30% there was clear representation of
197 one or more clusters in only one sample of the pair (e.g. P002 and P046, Fig. 3). These
198 exclusive clusters were found in both baseline and surgical samples of all three groups. In
199 these cases, we still found that at least one cluster, usually the one with mutations having
200 the highest VAFs in both samples, which was present in both samples of the pairs.

201

202 *Mutation detection in phase II*

203 The capture-probe panel from Phase I was used on additional samples from 28
204 patients (Fig. 1b) and 8 samples from Phase I where WES was unsuccessful, but enough DNA
205 was available. Sequencing of one sample from Phase I was unsuccessful. In concordance
206 with the analysis in Phase I, we excluded germline mutations based on their sequenced
207 matched blood. The median coverage for these samples was 91x (germline 103x, tumour
208 76x, Supplementary Data 1). One patient was excluded from further analysis because of a
209 different SNP profile (P085, Supplementary Fig. 9). The mutation count for the remaining 27
210 patients without prior WES discovery was inevitably much lower than for Phase I samples

(mean 6.4, median 5.0 mutations per patient, interquartile range, 3.0 to 6.0, Supplementary Fig. 10) as only few mutations in the Phase II were found outside the 77 breast cancer related genes. As for Phase I we only used tier 1 mutations for further analyses (Supplementary Data 3).

Frequently mutated genes

We combined the mutation data from Phase I and II to identify frequently mutated genes in the samples of the 86 patients in our dataset (Table 2). Six of the 77 breast cancer related genes were mutated in 10% or more of the patients. In decreasing frequency these were *PIK3CA* (37%), *TP53* (26%), *CDH1* (14%), *MLL3* (14%), *ABCA13* (12%) and *FLG* (10%). The top three genes are also the most frequently mutated genes in ER+, post-menopausal breast cancers in TCGA[8] (Supplementary Table 3). The frequency of mutations in *PIK3CA* and *CDH1* was very similar to the TCGA cohort, but the other four genes showed higher frequency in our data set, especially *ABCA13* with 12% compared with 4% in TCGA. We assessed whether good or poor responders were significantly associated with mutations in *ABCA13* or other frequently mutated genes, but we did not find such an association (6/27 *vs* 2/31, p=0.258, Fisher's exact test, not shown for other genes). Apart from the top three frequently mutated genes (*PIK3CA*, *TP53* and *CDH1*) only *GATA3, RYR2* and *MAP3K1* are mutated in more than 5% of patients in TCGA (9%, 6% and 9% of tumours, respectively). For these we found similar frequencies in our set (7% 6% and 5%, respectively). The most recurrent amino acid changes in our patients were H1047R (in 14 patients) followed by E545K (5 patients) in *PIK3CA*. For the majority of the frequently mutated genes missense was the most common amino acid change. Exceptions were *CDH1* with predominantly

frameshift mutations (12 frameshift, 1 missense and 1 nonsense) and *MLL3* with nonsense mutations (14 nonsense, 4 missense and 1 frameshift).

There was at least one mutation in a frequently mutated gene in 53 of the 77 pairs (Fig. 4). In all but 22 cases, the mutations in frequently mutated genes were identical for both samples of the pair giving a 54% concordance. In these pairs, 28 sites were identified as discordant, although 14 of these showed a measurable frequency (but not reaching statistically significance) in the other sample of the pair. The other discordant sites showed no frequency in the other samples of the pair, but all had a coverage >50x. The mutation status per patient (identical mutations or wild type) of the 6 frequently mutated genes was concordant in 71% of the complete set of 77 pairs. For individual genes, the concordance was higher for *PIK3CA* (3/27 discordant/concordant, 90%) and *TP53* (7/15, 68%) compared to the less frequently mutated genes *ABCA13* (6/2, 25%) and *FLG* (6/4, 40%). Also the VAF of mutations in *PIK3CA* (median baseline/surgery 30.3%/28.8%) and *TP53* (33.3%/33.1%) were generally higher than for *ABCA13* (15.5%/11.1%) and *FLG* (12.3%/13.5%), which were lower than the overall median of 25.7%.

Mutations in breast cancer driver genes listed by DriverDB[9] were found in 65 of the 77 sample pairs with a median of 2 driver gene mutations per sample (Supplementary Table 4). In 25 pairs all driver mutations were identified in both samples. Twelve pairs had none of their driver mutations shared, resulting in an overall concordance of 54%.

TP53 and HER2

Non-functional *TP53* can lead to DNA damage accumulation[10]. Therefore, we compared the mutational load of samples from Phase I by their *TP53*-mutation status and found a significantly higher load for mutated samples (median WT 37 *vs* mutant 64.5, p=0.017, Mann-Whitney test). For samples from Phase I the mutational load correlated weakly with Ki67 level at baseline (r=0.31, p=0.02 Spearman correlation), but a moderate correlation was found for the treated samples at surgery (r=0.40, p=0.01, Fig. 5a). Poor responders and *TP53* are both associated with higher mutational load: using the combined set of patients (Phase I and II), we hypothesized that poor responders were more likely to have a *TP53* mutation compared with good responders, but this hypothesis was rejected (10/23 *vs* 8/25, p=0.8, Fisher's exact test). However, we found a significantly higher Ki67 baseline level for *TP53* mutated samples (Supplementary Fig. 11) for both good (median WT 16.9 *vs.* mutated 36.7, p=0.020, Mann-Whitney test) and poor responders (median WT 15.9 *vs.* mutated 32.3, p=0.006). This difference was lost after treatment for the good, but persisted for poor responders (median WT 10.3 *vs.* mutated 28.7, p=0.011, Fig. 5b).

In HER2+ and HER2- tumours the median mutational load was 64 and 42, respectively (p=0.180, Mann-Whitney test). There was a higher than expected HER2+ rate amongst the control samples (35% in this dataset, expected rate in an ER+ population is ~10%[11]).

A significant decrease in the cellularity between baseline and surgery samples was found for good, but not poor responders or Controls (Supplementary Fig. 12) as reflected by the total number of cells per high-powered field in the Ki67 analysis. The type of biopsy taken at surgery (core-cut or resection) did not differ statistically between any responder

277 groups and did not explain differences in cellularity for good and poor responders

278 (Supplementary Fig. 13).

279

Discussion

Our primary goal was to identify DNA changes that relate significantly to the response of ER+ breast cancer to short-term oestrogen deprivation using AIs. Although the pre-surgical setting was ideal for this purpose, little is known about the true, as opposed to theoretical, impact of tissue heterogeneity on mutational profiling from the small tumour biopsies that are available for mutation profiling studies of clinical material. Our data on reproducibility are critical for a valid understanding of the current study and the many other studies of this type.

Very few data have been published on the genomic heterogeneity of multiple cores taken from the same breast tumour. The correlations of VAFs from two samples from five breast tumours reported by Ellis *et al*[6] (r=0.74 to r=0.94) were consistent with the majority of comparisons in the current analysis but in our larger set the correlations were much lower for some of our cases (Supplementary Fig. 5). Preliminary data was recently reported on 13 patients with multiple (7 to 17) spatially separated samples of primary breast cancer (ER+ and other types) – heterogeneity was observed within the samples even of cancer driver mutations[12]. Yates *et al*[13] reported heterogeneity in 8 out of 12 treatment-naive breast cancers based on 8 spatially separated biopsies from each tumour.

Most pairs in our study showed several clusters (potential subclones) present in both samples, but in about 30% of the cases we also found sample pairs with several clusters being exclusive to either sample and therefore spatially separated in the same tumour. However, these pairs shared at least one cluster, usually the one with the highest VAFs, indicative of a common founding clone with driving cancer mutations[14]. Although clusters exclusive to one sample were often present in a small proportion of sequenced cells, each subclone potentially has different adaptive properties and might become the dominant

clone due to selection from treatment[15,16]. Clusters disappearing or becoming more prominent in the treatment group could be indicative of such a selection. In our data it is unlikely that the exclusive clusters occur due to selection from AI treatment since we found exclusive clusters in the Control group as well and AI treatment had a very modest effect on cellularity in the 2 weeks of this study.

Reduced heterogeneity was found after 6 months of AI treatment[17]. In our data after much shorter time, we found that the number of mutations and the VAFs were slightly but statistically significantly lower in the surgical samples of treated compared with control patients, possibly indicating a modest treatment effect. Such a small effect was consistent with the slow rate of clinical response of tumours to endocrine therapy. This is dependent on cytostasis and not enhanced cell death such that tumour shrinkage is rarely apparent over a 2 week time period. In the good responder group, we noted that a minor loss of cellularity occurred over the 2-week period based on field counts of nuclei. Reduced cellularity could conceivably make the microdissection we carried out for all tissue sections prior to genomic analysis more difficult and thereby lead to greater non-malignant cell contamination potentially reducing the sensitivity to detect variants. These results are therefore consistent with the slightly decreased number of mutations in the surgical samples being at least in part an artefact of the lower malignant cell purity in the dissected material from the surgical samples. Given that the median loss between baseline and surgical samples from AI-treated patients was only 2 mutations we rationalized that surgical samples even from these were sufficiently unaffected by treatment to be acceptable as representative of the untreated state. Merging mutation data from baseline with surgical

samples including those from treated patients should provide more comprehensive information on the mutational landscape of a tumour than single cores.

Modest coverage for WES might have missed mutations with low VAF, especially mutations present at very low frequency in both samples of a pair. These mutations therefore could not be integrated in the panel and subsequently are missing in the final set of mutations and subclones. To maximise the number of mutations in the capture-panel, we used the union of several variant callers on the WES data to detect potential somatic mutations. Further, we included the entire coding sequencing of 77 breast cancer related genes in the panel to be able to detect mutations in these independent of the discovery step. We used the same capture-probe panel for additional samples in Phase II of this study. Unlike Phase I, the panel was not specifically designed to validate mutations found in the discovery stage. Therefore, in Phase II far fewer mutations per sample were found outside the 77 breast cancer related genes, emphasizing the individuality of the mutational profile of each breast cancer tumor[18]. For the combined set of samples from Phase I and II we therefore exclusively concentrated on the 77 breast cancer genes present on the targeted-panel and did not perform analyses based on mutation count or subclonality with these.

As expected, the most frequently mutated genes across the 86 patients were the breast cancer driver genes *PIK3CA* and *TP53*[6]. *CDH1*[19] and *MLL3*[20] are also frequently mutated genes known to be linked to breast cancer. The genes *FLG* and *ABCA13* are less studied, but *FLG* was shown to be amplified in a subset of breast cancers[21]. The frequency of patients with mutations in *ABCA13* was about three-fold higher compared to post-menopausal ER+ breast cancer tumours from TCGA[8]. A reason for this could be the selection of patients based on good and poor response; however, we did not find significant

differences between good and poor responders in terms of the mutation frequency of *ABCA13* or other frequently mutated genes. *GATA3* was not in the top list of frequently mutated genes, but its frequency was similar to that in TCGA. It was suggested that *GATA3* mutations might be a positive prediction marker for AI response based on Ki67 decline[6]. Our data cannot support this finding, but the statistical power with 6 mutated patients is low.

We saw low correlations for some samples based on the VAF values of all mutations in a sample pair. However, the mutation status of frequently mutated genes in the present data was found to be consistent within pairs in 76% of cases. Thus in a majority of cases the profile of mutations in the genes would be represented by one core. However, in about one in four patients this would not be the case and a single core-cut would have missed a potentially important gene mutation. We noted higher discordance and lower VAFs for mutations in less frequently mutated genes (*ABCA13* and *FLG*). This suggests that these mutations are subclonal, but might have important functions upon selective pressure. However, mutations at lower VAF are also more difficult to detect, which might in part explain the lower concordance for these mutations. We also analysed the concordance for the more numerous driver genes listed in DriverDB and we found a lower concordance of 54% between all pairs.

To study the impact of mutational profile on response to AI treatment, patients at the extreme ends of the Ki67 response spectrum were chosen as poor or good responders from the available patient sample set. Change in Ki67 after 2 weeks is a validated end-point for benefit from adjuvant endocrine therapy while the value of Ki67 after 2 weeks is prognostic for recurrence-free survival[22]. Ellis *et al*[6] related the mutational profile to resistance to AI in 77 patients using Ki67, defining resistance as on treatment Ki67 > 10%

372 irrespective of starting level. According to this definition four patients in our data set would
373 have been categorized as good responders despite exhibiting a minimal Ki67 decrease.
374 Nonetheless, there is generally good concordance between these two definitions and the
375 major conclusions on AI resistance from the current study and the Ellis study are similar.

376 We excluded 7% of patients who were categorized as poor responders according to
377 Ki67 decrease due to a lack of E2 suppression. It is not known whether this was due to poor
378 compliance or poor pharmacologic response but whichever is the case this highlights the
379 importance of measuring primary pharmacological response to avoid intensive molecular
380 investigation of tumours for mechanisms of resistance when the expected pharmacological
381 perturbation is absent.

382 The relatively low frequency of mutations in most genes in primary breast cancer
383 means that large studies are required to define reliable associations with
384 response/resistance to therapy even in pre-surgical studies such as POETIC where biological
385 response is measurable in all treated patients (in contrast to adjuvant therapy).
386 Nonetheless, we found a reduced suppression of Ki67 for *TP53*-mutated tumours within the
387 poor responder group, which supports the finding by Ellis *et al*[6] who reported a greater
388 suppression of Ki67 by letrozole in wild type than *TP53*-mutated tumours. This indicates at
389 least in part that *TP53* mutations are a marker for poor response to AI in addition to being a
390 marker for poorer outcome for ER+ breast cancer. We also found a significant association of
391 mutated *TP53* with increased mutational load. For *TP53* this is consistent with it being an
392 important DNA repair gene, malfunction of which may lead to general genomic instability
393 and an increase in mutations. The association of these factors with high mutational load was
394 recently reported by Haricharan *et al*[23].

It could be expected that poor responders to endocrine therapy might exhibit greater genomic heterogeneity given its potential to provide multiple pathways of resistance, a hypothesis supported by the larger number of mutations found in poor responders in this study. The clear presence of subclonality and multiple driver mutations in some of these early breast tumours does indicate the potential for some subclones to be selected preferentially during hormonal treatment and to drive the clinical regrowth of a partially responsive tumour. Identification of such subclones or mutations requires further studies on a later time point when the effect of treatment would be greater than that at two weeks.

In conclusion, this study demonstrates that multiple subclones are present even in early ER+ breast cancer. In most cases the subclones and their constituent mutations are represented in different core-cuts from the same tumour but in about 30% of the tumours mutations are exclusive to one of the core-cuts. Increased mutational load is associated with poorer antiproliferative response to AI possibly driven by mutations in *TP53*.

Methods

Patients and tissues

The design and goals of the POETIC trial (CRUK/07/015) have been published[2]. In brief, post-menopausal patients with primary ER and/or PgR-positive (according to local testing) breast cancer in over 120 centres across the UK were randomized 2:1 to receive or not receive an AI (anastrozole 1mg/d or letrozole 2.5mg/d) for a four-week period starting two weeks prior to surgery.

Core-cut biopsies (14-gauge) and either core-cuts or part of the excision sample were collected at baseline and surgery, respectively, and fixed in formalin. Additional core-cuts were collected into RNA*later* (Qiagen) at both time points. Whole blood was collected for germline DNA analysis, baseline and surgical plasma for estradiol analysis.

The trial was approved by the NRES Committee London - South East. All patients gave informed consent for DNA sequencing.

Biomarker analyses

Ki67% staining (MIB-1 clone code n. M7240, DAKO UK Ltd; working dilution 1:40) was the primary biomarker end-point for the POETIC trial and was centrally analysed on all formalin-fixed samples using a single protocol (either core-cut in FFPE or excision specimens in FFPE) as previously described[22]. All staining was performed on a Dako autostainer using strict adherence to a single staining protocol. Hematoxylin and eosin (H&E) staining was used to exclude samples with low tumour purity (<40%).

HER2 status was measured locally using immunohistochemistry and/or in situ hybridization[24]. Biomarker results are shown in Supplementary Table 5.

ER expression of baseline specimens was measured by immunohistochemistry (6F11 clone code n. NCL-L-ER-6F11, Leica Biosystems Ltd; working dilution 1:50) on formalin-fixed samples[25]. Patients were excluded from this sub-study if they were described as ER negative (< 1% positive staining of tumour nuclei).

Cellularity was measured by 10x10mm eye-piece graticule with 40x objective graticule. Nuclei were counted within the grid of at least 5 fields and the mean values from these measurements were used.

Patients with unsuppressed estradiol upon treatment were excluded.

Sample selection

In Phase I samples were selected with the aim of having equal numbers of control patients, definite poor responders defined as having a Ki67 decrease of <60% between baseline and surgery and good responders with >75% Ki67 decrease. The definition of good responders was selected as being above the mean Ki67 reduction to anastrozole after two weeks[26]. Patients with Ki67 decrease between 60 and 75% were excluded to create an efficient design that focused on the extremes of the range of Ki67 responses. Treated patients not showing suppressed post-menopausal levels of plasma estradiol and those with central ER <1% were excluded. For Phase II only treated samples were selected.

DNA extraction

DNA was extracted from RNA*later* preserved diagnostic (baseline) and surgical (surgery) 14-gauge core-cut samples and peripheral blood.

At least eight unstained eight-micron sections were taken from core-cuts embedded in OCT (Cryo-M-Bed, Bright Instruments, UK). Sections were stained with Nuclear Fast Red (0.1% (w/v)) and when necessary needle microdissection was used to achieve >60% pure tumor cells) using an adjacent haematoxylin and eosin-stained section as a guide. DNA was extracted from the sections using the DNeasy Tissue and Blood Kit (Qiagen) and from blood using the EZ1 system (LifeTechnologies).

Exome sequencing for discovery

Cavitation (adaptive focused acoustics, Covaris) was used to fragment the samples. The automated libraries were generated with in-house Illumina kits at Washington University, MO, with reagents supplied by NEB and indexed via PCR. LucigenDNATerminator Kit (end repair), NEB Klenow (Adenylation), NEB Quick Ligase (ligation, Illumina's Multiplexing Adapters), NEB Phusion (PCR enrichment, libraries were indexed via PCR (PCR1.0, PCR2.0 and index primers), AMPure beads were used for enzymatic purification and size selection). Manual libraries were generated with KAPA Library Preparation with standard PCR library amplification (KK8201) and libraries were indexed during ligation with TruSeq LT adaptors. LabChip GX was used for library quantitation as well as quality control. Size selection was conducted using AMPure beads. 10 libraries were pooled pre-capture. Each library pool was captured using NimblegenSeqCap EZ Human Exome Library v3 (with requisite SeqCap EZ hybridisation and wash kits) and sequenced on two lanes of the IlluminaHiSeq 2000 with v3 chemistry (2x100bp).

Sequence data was aligned to reference sequence build GRCh37-lite-build37 using bwa version 0.5.9[27] (params: -t 4 -q 5) then merged using picard version 1.46 (http://picard.sourceforge.net), then deduplicated using picard version 1.46.

SNVs were detected using the union of three callers: 1) samtools version r963[28] (params: -A -B) intersected with Somatic Sniper version 1.0.2[29] (params: -F vcf -q 1 -Q 15) and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15 --min-mapping-quality 40 --min-somatic-score 40) 2) VarScan version 2.2.6[30] filtered by varscan-high-confidence filter version v1 and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15 --min-mapping-quality 40 --min-somatic-score 40), and 3) Strelka version 0.4.6.2[31] (params: isSkipDepthFilters = 1).

InDels were detected using the union of 4 callers: 1) GATK somatic-indel version 5336[32] filtered by false-indel version v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15), 2) pindel version 0.5[33] filtered with pindel false-positive and vaf filters (params: --variant-freq-cutoff=0.2), 3) VarScan version 2.2.6 [30] filtered by varscan-high-confidence-indel version v1 then false-indel version v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15), and 4) Strelka version 0.4.6.2[31] (params: isSkipDepthFilters = 1).

Targeted sequencing for validation

All of the variants (n=6,910) identified in the discovery set excluding those in low coverage samples were chosen for validation at greater depth as well as exons of a set of 77

499 breast cancer related genes of interest (Supplementary Table 2). Probes were designed to
500 target the variants within 13,372 regions of 6,737 genes covering a total of 2,645,703bp.

501 Reads were aligned as described above for exome-sequencing. SNVs were detected
502 using VarScan version 2.2.6 (with parameters --min-var-freq 0.08 --p-value 0.10 --somatic-p-
503 value 0.01 --validation) and filtered by Varscan-high-confidence version v1, then false-
504 positive version v1 (with parameters: --bam-readcount-version 0.4 --bam-readcount-min-
505 base-quality 15).

506 InDels were detected using the union of 3 callers: GATK somatic-indel version 5336,
507 pindel version 0.5 (filtered by pindel-somatic-calls version v1, then pindel-read-support
508 version v1) and VarScan version 2.2.6 (filtered by varscan-high-confidence-indel version v1,
509 then false-indel version, with parameters: --bam-readcount-version 0.4 --bam-readcount-
510 min-base-quality 15).

511 In addition to using matched normal for germline detection, sites that were present
512 in at least 0.1% of the general population according to the 1000 Genomes Project[34] or NHLBI
513 GO Exome Sequencing Project were removed from further analysis.

514 All somatic events from re-sequencing were manually reviewed using IGV[35].

515

516 *SNP profile*

517 Samples were confirmed as being derived from the same patient by correlation of
518 single nucleotide polymorphisms (SNPs, Supplementary Fig. 9) based on the sequencing
519 data. Therefore, all samples have been profiled based on over 500 SNPs from dbSNP version
520 138 within the area of the capture panel. The genotypes at genomic position were derived

using samtools[28]. SNPRelate[36] was used to cluster the samples and generate the dendrogram using default parameters.

Subclonal analysis

Clonal architecture was inferred using SciClone version 1.0.4[7] (params: minimumDepth=50) using copy number and LOH calls derived from Varscan (params: loh-cutoff=0.95, min-loh-probes=10, min-mapping-quality=10, min-coverage=20. min-segment-size=25, max-segment-size=100, undo.sd=4). Samples with low mutation count failed clustering and were excluded from the analysis. SciClone plots were annotated with frequently mutated genes from Supplementary Table 6.

Comparison to known driver genes

Validated mutations in the baseline and surgery samples were compared to known driver genes in the DriverDB[9] database. Therefore, "breast" tissue was selected as cancer type and all genes identified by at least 2 tools were downloaded from the website.

Estimating tumour purity based on WES

For estimating cellularity based on whole exome sequencing data Sequenza v2.1.1[37] was used. The algorithm was applied for each tumour sample and its matched blood sample. In brief, it first detects germline mutations in the normal sample and then calculates the VAF at the same position in the tumour sample. In the second step the tumour versus normal depth ratio is calculated with GC-content normalization and allele-specific

segmentation is performed. Based on a probabilistic model applied to segmented data Sequenza calculates possible solutions for cellularity and ploidy of the tumour. The default settings were used for all steps, the cellularity with the highest probability was reported.

Statistical analysis

Unpaired and nonparametric Mann-Whitney test was used to test the differences of mutation counts between groups. The Wilcoxon signed-rank test was used to test for differences in the mutation counts of paired samples between baseline and surgery and to compare the variant allele fractions (VAFs) between baseline and surgery samples. The associations of *TP53* mutation status and HER2 status between the groups were analysed using Fisher's exact test. Reported p-values are two-sided and unadjusted; p-value < 0.05 is considered to be significant in this study. The statistical analyses were conducted in GraphPad Prism 6 (Graphpad Software Inc.) and R[38].

Data availability

The sequencing data that support this study have been deposited in the European Genome-phenome Archive (EGA) database under accession code EGAS00001001940. The remaining data is available in the Article or Supplementary files or available from the authors upon request.

References

1 Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine* **366**, 883-892, doi:10.1056/NEJMoa1113205 (2012).

2 Dowsett, M. *et al.* Endocrine therapy, new biologicals, and new study designs for presurgical studies in breast cancer. *Journal of the National Cancer Institute. Monographs* **2011**, 120-123, doi:10.1093/jncimonographs/lgr034 (2011).

3 Cancer.net. *Breast Cancer Guide*, <http://www.cancer.net/cancer-types/breast-cancer/overview> (2015).

4 Early Breast Cancer Trialists' Collaborative, G. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet*, doi:10.1016/S0140-6736(15)61074-1 (2015).

5 Smith, I. E. & Dowsett, M. Aromatase inhibitors in breast cancer. *The New England journal of medicine* **348**, 2431-2442, doi:10.1056/NEJMra023246 (2003).

6 Ellis, M. J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353-360, doi:10.1038/nature11143 (2012).

7 Miller, C. A. *et al.* SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution. *PLoS computational biology* **10**, e1003665, doi:10.1371/journal.pcbi.1003665 (2014).

8 Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70, doi:10.1038/nature11412 (2012).

9 Cheng, W. C. *et al.* DriverDB: an exome sequencing database for cancer driver gene identification. *Nucleic acids research* **42**, D1048-1054, doi:10.1093/nar/gkt1025 (2014).

10 Hanel, W. & Moll, U. M. Links between mutant p53 and genomic instability. *J Cell Biochem* **113**, 433-439, doi:Doi 10.1002/Jcb.23400 (2012).

11 Anderson, H. *et al.* Relationship between estrogen receptor, progesterone receptor, HER-2 and Ki67 expression and efficacy of aromatase inhibitors in advanced breast cancer. *Ann Oncol* **22**, 1770-1776, doi:DOI 10.1093/annonc/mdq700 (2011).

12 Connolly, L. *et al.* Genomic heterogeneity in primary breast cancer: Clinical implications. *J. Clin. Oncol.* **35:5s**, (suppl; abstr 11004) (2014).

13 Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21**, 751-759, doi:10.1038/nm.3886 (2015).

14 Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339, doi:10.1038/nature12634 (2013).

15 Keats, J. J. *et al.* Clonal competition with alternating dominance in multiple myeloma. *Blood* **120**, 1067-1076, doi:10.1182/blood-2012-01-405985 (2012).

16 Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506-510, doi:10.1038/nature10738 (2012).

17 Quenel-Tueux, N. *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, doi:10.1038/bjc.2015.247 (2015).

18 Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404, doi:10.1038/nature11017 (2012).

19 Lei, H. *et al.* CDH1 mutations are present in both ductal and lobular breast cancer, but promoter allelic variants show no detectable breast cancer risk. *Int. J. Cancer* **98**, 199-204 (2002).

20 Wang, X. X. *et al.* Somatic mutations of the mixed-lineage leukemia 3 (MLL3) gene in primary breast cancers. *Pathol. Oncol. Res.* **17**, 429-433, doi:10.1007/s12253-010-9316-0 (2011).

21 Adnane, J. *et al.* BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* **6**, 659-663 (1991).

22 Dowsett, M. *et al.* Prognostic value of Ki67 expression after short-term presurgical endocrine therapy for primary breast cancer. *Journal of the National Cancer Institute* **99**, 167-170, doi:10.1093/jnci/djk020 (2007).
23 Haricharan, S., Bainbridge, M. N., Scheet, P. & Brown, P. H. Somatic mutation load of estrogen receptor-positive breast tumors predicts overall survival: an analysis of genome sequence data. *Breast cancer research and treatment* **146**, 211-220, doi:10.1007/s10549-014-2991-x (2014).
24 Wolff, A. C. *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* **31**, 3997-4013, doi:10.1200/JCO.2013.50.9984 (2013).
25 Hammond, M. E. *et al.* American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch. Pathol. Lab. Med.* **134**, e48-72, doi:10.1043/1543-2165-134.7.e48 (2010).
26 Dowsett, M. *et al.* Short-term changes in Ki-67 during neoadjuvant treatment of primary breast cancer with anastrozole or tamoxifen alone or combined correlate with recurrence-free survival. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 951s-958s (2005).
27 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
28 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
29 Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28**, 311-317, doi:10.1093/bioinformatics/btr665 (2012).
30 Koboldt, D. C. *et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283-2285, doi:10.1093/bioinformatics/btp373 (2009).
31 Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817, doi:10.1093/bioinformatics/bts271 (2012).
32 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
33 Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871, doi:10.1093/bioinformatics/btp394 (2009).
34 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).
35 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
36 Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-3328, doi:10.1093/bioinformatics/bts606 (2012).
37 Favero, F. *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**, 64-70, doi:10.1093/annonc/mdu479 (2015).
38 R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2010).

665

Acknowledgments

Lila Zabaglo, Elizabeth Folkerd, Margaret Hills, Simone Detre and Maria Afentakis provided excellent laboratory support. We thank all participating patients and staff at POETIC centres, and ICR-CTSU who contributed to trial management and data and sample collection, and the independent data monitoring committee and trial steering committee for their oversight of the trial. This work was supported by the Mary-Jean Mitchell Green Foundation, Breast Cancer Now and the NIHR RM/ICR Biomedical Research Centre. The POETIC trial (C1491/A8671/CRUK/07/015, C1491/A15955, C406/A8962), from which samples were obtained for this study, was supported by Cancer Research UK (CRUK/07/015) as is ICR-CTSU through its core programme grant.

POETIC was registered with ISRCTN registry (ISRCTN63882543).

Author Contributions

PG and MD wrote the manuscript. CVS, PG, LAM, ERM, JR, JB, IS and MD contributed to the overall design of this study. Data was collected by CVS, ERM, JR, JB, AG, MG, ELK, KS, AE, MS, CH, SM, NB, AS, WM and MD. Data was analysed and interpreted by PG, CVS, QG, LAM, TL, CAM, ELK, CL, ERM, JB, IS and MD. All authors have read and approved the final version of the manuscript.

Competing Financial Interests

JR is employed by Oncimmune. CH is a paid consultant to Genomic Health, MD is a paid consultant to Pfizer and NanoString. JR holds stock in Oncimmune and FaHRAS. CH received

travel expenses from Novartis, JR from Oncimmune, AstraZeneca, Bayer, Novartis and Syndax. JR received honoraria from AstraZeneca, Bayer and Amgen. JR receives research funding from Oncimmune, MD from Novartis, AstraZeneca and Pfizer. JR holds patents from Oncimmune and is in the speakers' bureau of AstraZeneca. The remaining authors declare no competing financial interests.

POETIC trial members

Stephen Johnston[4,13], Radha Todd[14], Kieran Horgan[15], Stephen Chan[16], Simon DH Holt[17], Marina Parton[4], Ian Laidlaw[18], Jayant S Vaidya19, Tracey Irvine[20], Fiona Hoar[21], Ilyas Khattak[22], Ashutosh Kothari[23], Lucy Brazil[23], Nicholas Gallegos[24], Duncan Wheatley[25], Tayo Johnson[26], Geoffrey Sparrow[27], Serena Ledwidge[28], Caroline Mortimer[29], Marcus Ornstein[30], Douglas Ferguson[31], Douglas Adamson[32], Ramsey Cutress[33], Richard Johnson[34], Clare Crowley[35], Zoe Winters[36], Hisham Hamed[37], Russell Burcombe[38,39], Susan Cleator[40,461], Muireann Kelleher[42], Jonathan Roberts[43], Sarah Vesty[44], Maher Hadaki[45], Mary Quigley[46], Julie Doughty[47], Siobhan Laws[48], Seema Seetharam[49], Amanda Thorne[50], Peter Donnelly[51]

[13]Royal Marsden Hospital, Downs Road, Sutton, SM2 5PT

[14]Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP

[15]St James's University Hospital, Beckett Street, Leeds, LS9 7TF

[16]Nottingham City Hospital, Hucknall Road, Nottingham, NG5 1PB

[17]Prince Phillip Hospital, Dafen, Llanelli, SA14 8QF

[18]Frimley Park Hospital, Portsmouth Road, Frimley, GU16 7UJ

[19]Whittington Hospital NHS Trust, Highgate Hill, London, N19 5NF

[20]Royal Surrey County Hospital, Egerton Road, Guildford, GU2 7XX

[21]City Hospital, Dudley Road, Birmingham, B18 7QH

[22]Ysbyty Gwynedd, Penrhosgarnedd, Bangor, LL57 2PW

[23]Guy's Hospital, Great Maze Pond, London, SE1 9RT

[24]Weston General Hospital, Grange Road, Uphill, Weston-Super-Mare, BS23 4TQ

[25]Royal Cornwall Hospital, London Road, Treliske, Truro, TR1 3LJ

[26]St Peter's Hospital, Guildford Road, Chertsey, KT16 0PZ

[27]Yeovil District Hospital, Higher Kingston, Yeovil, BA21 4AT

[28]St Bartholomew's Hospital, West Smithfield, London, EC1A 7BE

[29]Ipswich Hospital, Heath Road, Ipswich, IP4 5PD

[30]Homerton University Hospital, Homerton Row, London, E9 6SR

[31]Royal Devon and Exeter Hospital, Barrack Road, Exeter, EX2 5DW

[32]Ninewells Hospital, Dundee, DD1 9SY

[33]Southampton General Hospital, Tremona Road, Southampton, SO16 6YD

[34]Neath Port Talbot Hospital, Baglan Way, Port Talbot, SA12 7BX

[35]Salisbury District Hospital, Odstock Road, Salisbury, SP2 8BJ

[36]Bristol Royal Infirmary, Marlborough Street, Bristol, BS2 8HW

[37]Guy's Hospital, Great Maze Pond, London, SE1 9RT

[38]Maidstone Hospital, Hermitage Lane, Barming, Maidstone, ME16 9QQ

[39]Tunbridge Wells Hospital, Tonbridge Road, Tunbridge Wells, TN2 4QL

[40]St Mary's Hospital, 41 Praed Street, London, W2 1NY

[42]St George's Hospital, Blackshaw Road, London, SW17 0QT

[43]King's College Hospital, Denmark Hill, London, SE5 9RS

[4] General Hospital, Sandford Road, Cheltenham, GL53 7AN

[45]Maidstone Hospital, Hermitage Lane, Barming, Maidstone, ME16 9QQ

[46]Queen's Hospital, Rom Valley Way, Romford, Essex, RM7 0AG

[41]Charing Cross Hospital, Fulham Palace Road, London, W6 8RF

[47]Western Infirmary, Dumbarton Road, Glasgow, G11 6NT

[48]Royal Hampshire County Hospital, Romsey Road, Winchester, SO22 5DG

[49]Darent Valley Hospital, Darenth Wood Road, Dartford, DA2 8DA

[50]Musgrove Park Hospital, Taunton & Somerset NHS Foundation Trust, Taunton, TA1 5DA

[51]Torbay District General Hospital, Newton Road, Lawes Bridge, Torquay, TQ2 7AA

Figure legends:

Fig. 1. CONSORT diagram and mutational landscape.

Samples were selected in two phases using the same quality criteria. Samples in Phase I (**a**) underwent whole exome sequencing (WES) at low coverage for mutation detection followed by capture-probe sequencing for validation. Our goal was to select the same number of Controls, good responders and poor responders, but due to the availability of samples and exclusion criteria, we were not able to identify 20 poor responders, instead 15 and 25 poor responders entered the analysis. In Phase II (**b**) samples which failed WES (not shown, see Supplementary Fig. 1) and samples from additional patients without prior WES were sequenced with the same capture-probe panel as in Phase I. To balance the number of patients in the responder groups, preferentially poor responders were added. When samples from Phase I and II combined, a total of 86 patients entered the downstream analysis, of which 77 are paired samples (see also Table 2). CONSORT diagram is simplified; a more detailed version can be found in Supplementary Fig. 1. (**c**) Mutation type of all validated mutations in the exome of 59 patients from Phase I and (**d**) number of mutations in each patient by responder groups. Identical mutations found in the baseline and surgery sample of the same patients appear once in this Figure only.

Fig. 2. Differences of mutation counts and treatment effects.

Analysis on the mutation load of samples with exome-wide mutation profile from Phase I. (**a-b**) Poor responder showed significantly more mutations than good responder on baseline (B) and surgery (S). (**c**) Also the number of mutations on a per-patient basis (mutations from B and S samples combined, counting identical mutations once only) was significantly higher

768 in poor responders. Median and interquartile ranges are shown as bars. (**d**) No significant
769 difference between the B and S mutation counts within responder groups between each of
770 the 49 paired samples. (**e**) Good and poor responders showed a significant, but low
771 reduction of the mean variant allele fractions (VAFs) of single nucleotide variants between B
772 and S. Whiskers show 95% confidence interval. Significance was tested by Mann-Whitney
773 test.

774

775 **Fig. 3.** Intra-tumour heterogeneity.

776
777 Five examples with clear intra-tumour heterogeneity are shown (see Supplementary Fig. 6-8
778 for plots of all samples). Some patients had clusters present in both samples (P007, P014
779 and P039), while others had several clusters that were found in either the baseline or
780 surgery sample (P002 and P046). The variant allele fractions of mutations are shown. Whole
781 exome sequencing was used for copy-number assessment and only mutations in copy
782 number neutral regions were plotted. Colours indicate assigned clusters by SciClone
783 (Methods). Cancer related genes listed in Supplementary Table 4 are labelled in the plots.

784

785 **Fig. 4.** Frequently mutated genes

786 Sample matrix for genes with mutations in 10% or more of the patients. All 163 tumour
787 samples from 86 patients (including 77 pairs) with baseline (B) and surgery (S) sample are
788 shown. For Phase I samples the bottom row shows if the sample successfully underwent
789 whole exome sequencing and therefore mutations identified in this sample were added to
790 the capture-probe panel. The *TP53* mutation of P038 and one of each mutation of *MLL3* for

P046 were not identical between B and S. The overall concordance between B and S samples of patients was 71%.

Fig. 5. Relation of mutations to Ki67

(**a**) Correlation of mutations counts to the Ki67 level was highest for treated samples at surgery. (**b**) On the combined set from Phase I and Phase II the Ki67 level of poor responders was significantly higher for patients with mutated *TP53* (mut) than wild-type *TP53* (WT). This was not seen for good responders, although Ki67 level for *TP53* mutated patients was higher on baseline (Supplementary Fig. 11). Significance tested by Mann-Whitney test, red lines show median and interquartile ranges.

Table 1. Clinical data summary of all 86 patients in this study. Patient's demographics are separated by poor responder, good responder and Control. All analysis based on somatic mutations within 77 breast cancer related genes are conducted on this set of patients combined from Phase I and II (no analysis was conducted on Phase II samples only). Analyses on the exome-wide mutational load were performed on samples from Phase I with whole exome sequencing. The demographics of these patients only are shown in Supplementary Table 1.

	Response group					
	Poor (n=33)		Good (n=33)		Control (n=20)	
	n	%	n	%	n	%
PgR status						
Positive	20	60.6	26	78.8	16	80.0
Negative	7	21.2	5	15.2	3	15.0
Not known	6	18.2	2	6.1	1	5.0
Histological subtype						
Ductal	27	81.8	24	72.7	17	85.0
Lobular	3	9.1	5	15.2	2	10.0
Mucinous	1	3.0	1	3.0	0	0.0
Mixed ductal and lobular	2	6.1	1	3.0	1	5.0
Not known	0	0.0	2	6.1	0	0.0
Pre-treatment tumour grade						
G1	2	6.1	0	0.0	3	15.0
G2	14	42.4	22	66.7	9	45.0
G3	10	30.3	4	12.1	5	25.0
Not known	7	21.2	7	21.2	3	15.0
No. of involved lymph nodes						
N0	20	60.6	20	60.6	11	55.0
N1-3	7	21.2	11	33.3	6	30.0
N4+	6	18.2	2	6.1	3	15.0
HER2 status						
Negative	25	75.8	32	97.0	13	65.0
Positive	8	24.2	1	3.0	7	35.0
Pre-treatment tumour size (cm)						
<2	12	36.4	11	33.3	7	35.0
2-5	19	57.6	22	66.7	12	60.0
>5	2	6.1	0	0.0	1	5.0
Surgery tumour size (cm)						
<2	12	36.4	13	39.4	8	40.0
2-5	20	60.6	20	60.6	10	50.0
>5	1	3.0	0	0.0	2	10.0
	Median	**IQR**	**Median**	**IQR**	**Median**	**IQR**
Age at randomization (years)	70	61 - 78	74	62 – 82	70	59 - 76
Time from randomization to surgery (days)	19	15 - 23	17	15 – 19	18.5	14 - 23.5

Table 2. Summary of available samples in this study. Analyses on the mutational load and the mutational clusters were performed on Phase I samples with exome sequencing available. Analyses based on 77 breast cancer related genes were performed on the combined set of Phase I and II samples. As indicated in the text, some analyses were performed on patients with paired baseline and surgery samples only.

Summary	Total	Good	Poor	Control
Phase I				
Samples with mutations discovered by exome-	108	45	26	36
Patients with paired exome-sequencing	49	21	11	17
Patients with exome-seq from either sample	59	24	15	20
Patients with exome-sequencing in baseline	52	22	12	18
Patients with exome-sequencing in surgery	56	23	14	19
Phase I and II				
Samples with targeted sequencing	163	64	59	40
Patients with targeted sequencing from either sample	86	33	33	20
Patients with paired targeted-sequencing	77	31	26	20
Patients with targeted sequencing in baseline	84	32	32	20
Patients with targeted sequencing in surgery	79	32	27	20

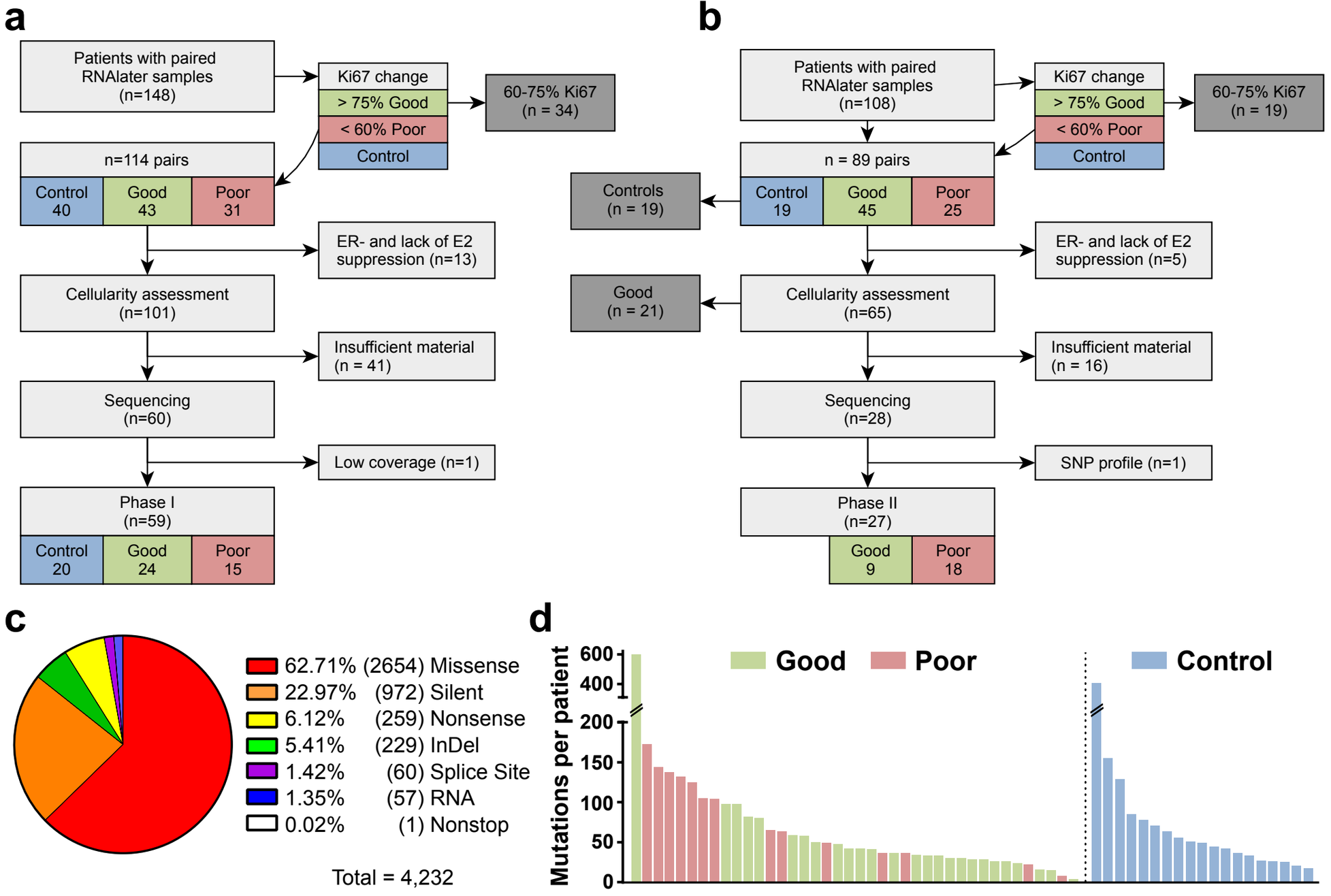

Fig. 1. CONSORT diagram and mutational landscape.
Samples were selected in two phases using the same quality criteria. Samples in Phase I (a) underwent whole exome sequencing (WES) at low coverage for mutation detection followed by capture-probe sequencing for validation. Our goal was to select the same number of Controls, good responders and poor responders, but due to the availability of samples and exclusion criteria, we were not able to identify 20 poor responders, instead 15 and 25 poor responders entered the analysis. In Phase II (b) samples which failed WES (not shown, see Supplementary Fig. 1) and samples from additional patients without prior WES were sequenced with the same capture-probe panel as in Phase I. To balance the number of patients in the responder groups, preferentially poor responders were added. When samples from Phase I and II combined, a total of 86 patients entered the downstream analysis, of which 77 are paired samples (see also Table 2). CONSORT diagram is simplified, a more detailed version can be found in Supplementary Fig. 1. (c) Mutation type of all validated mutations in the exome of 59 patients from Phase I and (d) number of mutations in each patient by responder groups. Identical mutations found in the baseline and surgery sample of the same patients appear once in this Figure only.

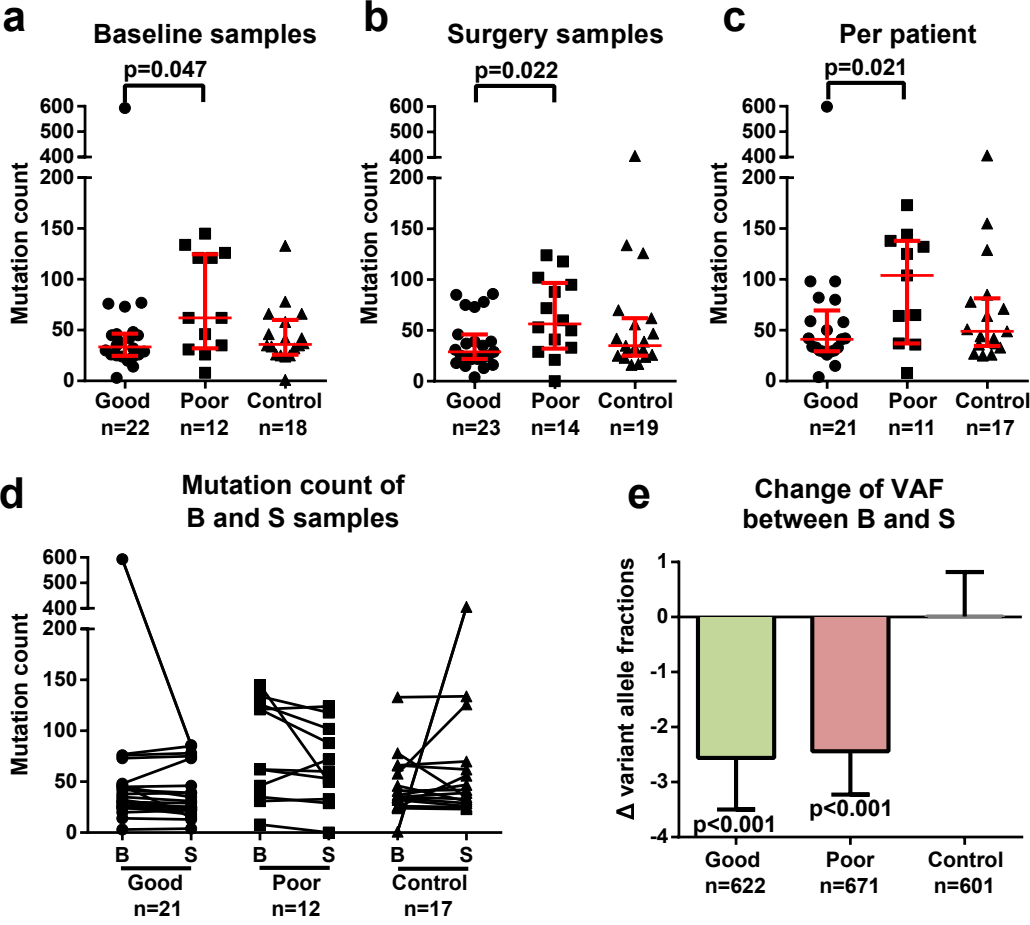

Fig. 2. Differences of mutation counts and treatment effects. Analysis on the mutation load of samples with exome-wide mutation profile from Phase I. (a-b) Poor responder showed significantly more mutations than good responder on baseline (B) and surgery (S). (c) Also the number of mutations on a per-patient basis (mutations from B and S samples combined, counting identical mutations once only) was significantly higher in poor responders. Median and interquartile ranges are shown as bars. (d) No significant difference between the B and S mutation counts within responder groups between each of the 49 paired samples. Red lines show median and interquartile ranges. (e) Good and poor responders showed a significant, but low reduction of the mean variant allele fractions (VAFs) of single nucleotide variants between B and S. Whiskers show 95% confidence interval. Significance was tested by Mann-Whitney test.

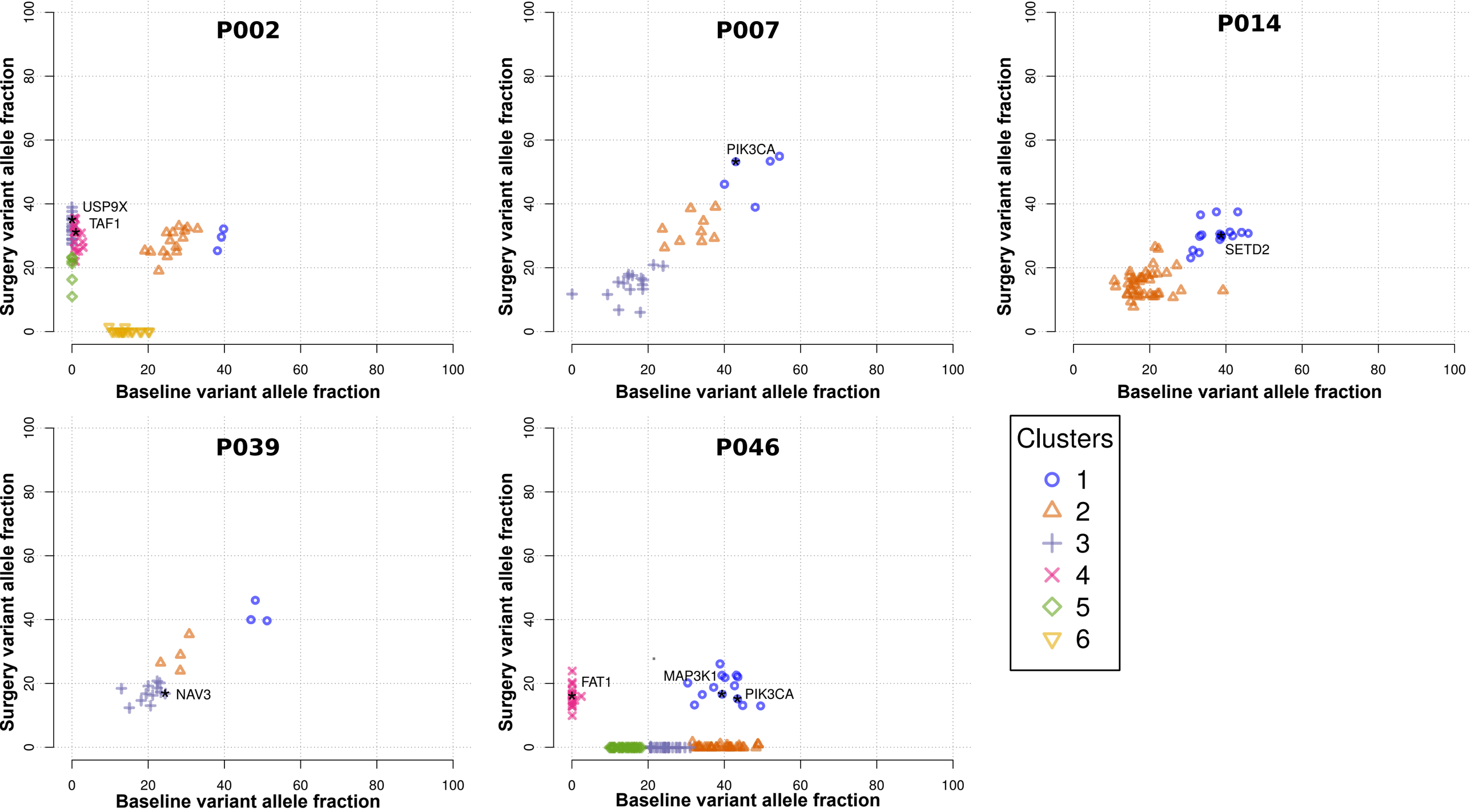

Fig. 3. Intra-tumour heterogeneity.
Five examples with clear intra-tumour heterogeneity are shown (see Supplementary Fig. 6-8 for plots of all samples). Some patients had clusters present in both samples (P007, P014 and P039), while others had several clusters that were found in either the baseline or surgery sample (P002 and P046). The variant allele fractions of mutations are shown. Whole exome sequencing was used for copy-number assessment and only mutations in copy number neutral regions were plotted. Colours indicate assigned clusters by SciClone (see Methods). Cancer related genes listed in Supplementary Table 4 are labelled in the plots.

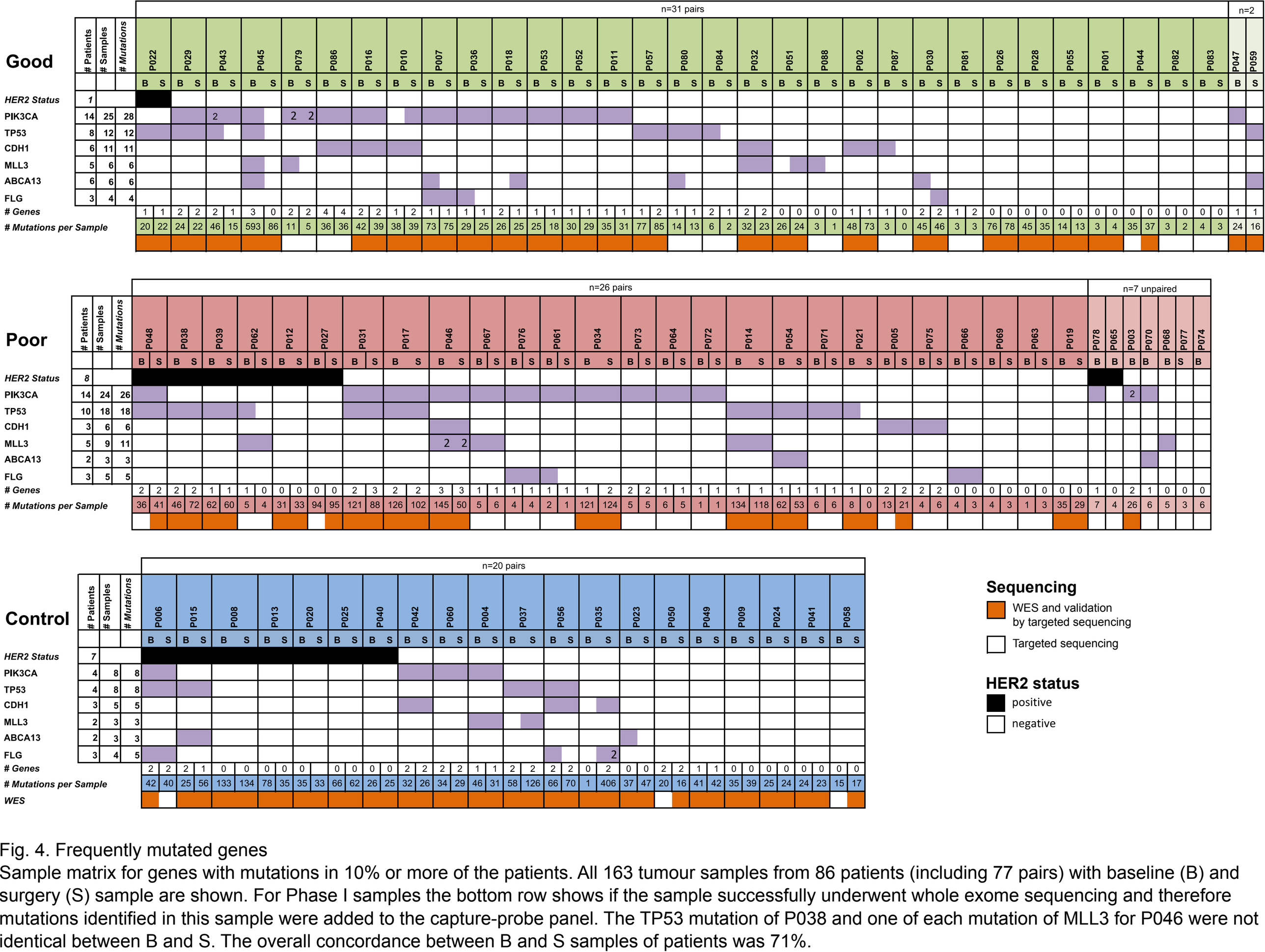

Fig. 4. Frequently mutated genes

Sample matrix for genes with mutations in 10% or more of the patients. All 163 tumour samples from 86 patients (including 77 pairs) with baseline (B) and surgery (S) sample are shown. For Phase I samples the bottom row shows if the sample successfully underwent whole exome sequencing and therefore mutations identified in this sample were added to the capture-probe panel. The TP53 mutation of P038 and one of each mutation of MLL3 for P046 were not identical between B and S. The overall concordance between B and S samples of patients was 71%.

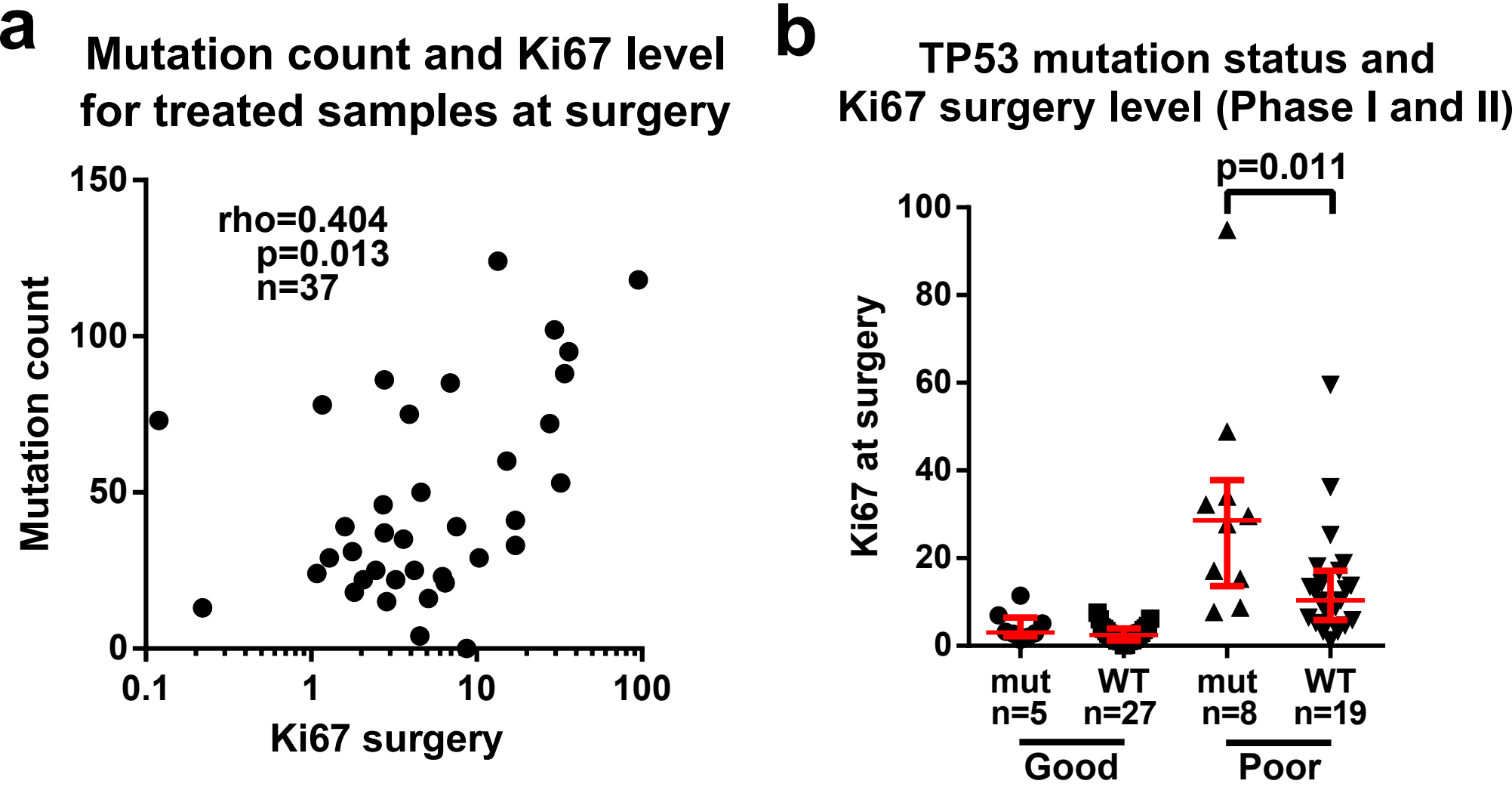

Fig. 5. Relation of mutations to Ki67
(a) Correlation of mutations counts to the Ki67 level was highest for treated samples at surgery (Spearman correlation). (b) On the combined set from Phase I and Phase II the Ki67 level of poor responders was significantly higher for patients with mutated TP53 (mut) than wild-type TP53 (WT). This was not seen for good responders, although Ki67 level for TP53 mutated patients was higher on baseline (Supplementary Fig. 11). Significance tested by Mann-Whitney test, red lines show median and interquartile ranges.

Table 1.

	Response group					
	Poor (n=33)		Good (n=33)		Control (n=20)	
	n	%	n	%	n	%
PgR status						
Positive	20	60.6	26	78.8	16	80
Negative	7	21.2	5	15.2	3	15
Not known	6	18.2	2	6.1	1	5
Histological subtype						
Ductal	27	81.8	24	72.7	17	85
Lobular	3	9.1	5	15.2	2	10
Mucinous	1	3	1	3	0	0
Mixed ductal and lobular	2	6.1	1	3	1	5
Not known	0	0	2	6.1	0	0
Pre-treatment tumour grade						
G1	2	6.1	0	0	3	15
G2	14	42.4	22	66.7	9	45
G3	10	30.3	4	12.1	5	25
Not known	7	21.2	7	21.2	3	15
No. of involved lymph nodes						
N0	20	60.6	20	60.6	11	55
N1-3	7	21.2	11	33.3	6	30
N4+	6	18.2	2	6.1	3	15
HER2 status						
Negative	25	75.8	32	97	13	65
Positive	8	24.2	1	3	7	35
Pre-treatment tumour size (cm)						
<2	12	36.4	11	33.3	7	35
2-May	19	57.6	22	66.7	12	60
>5	2	6.1	0	0	1	5
Surgery tumour size (cm)						
<2	12	36.4	13	39.4	8	40
2-May	20	60.6	20	60.6	10	50
>5	1	3	0	0	2	10
	Median	**IQR**	**Median**	**IQR**	**Median**	**IQR**
Age at randomization (years)	70	61 - 78	74	62 – 82	70	59 - 76
Time from randomization to su	19	15 - 23	17	15 – 19	18.5	14 - 23.5

Table 2.

Summary	Total	Good	Poor	Control
Phase I				
Samples with mutations discovered by exome-sequenci	108	45	26	36
Patients with paired exome-sequencing	49	21	11	17
Patients with exome-seq from either sample	59	24	15	20
Patients with exome-sequencing in baseline	52	22	12	18
Patients with exome-sequencing in surgery	56	23	14	19
Phase I and II				
Samples with targeted sequencing	163	64	59	40
Patients with targeted sequencing from either sample	86	33	33	20
Patients with paired targeted-sequencing	77	31	26	20
Patients with targeted sequencing in baseline	84	32	32	20
Patients with targeted sequencing in surgery	79	32	27	20