



HAL
open science

Validation of a yeast functional assay for p53 mutations using clonal sequencing.

Richard Iggo, Justine Rudewicz, Elodie Monceau, Nicolas Sevenet, Jonas Bergh, Tobias Sjoblom, Hervé Bonnefoi

► To cite this version:

Richard Iggo, Justine Rudewicz, Elodie Monceau, Nicolas Sevenet, Jonas Bergh, et al.. Validation of a yeast functional assay for p53 mutations using clonal sequencing.. *Journal of Pathology*, 2013, 231 (4), pp.441-8. 10.1002/path.4243 . hal-01070094

HAL Id: hal-01070094

<https://hal.science/hal-01070094>

Submitted on 22 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Validation of a yeast functional assay for *p53* mutations using clonal sequencing

Richard Iggo,^{1*} Justine Rudewicz,¹ Elodie Monceau,¹ Nicolas Sevenet,¹ Jonas Bergh,² Tobias Sjoblom³ and Hervé Bonnefoi¹

¹ INSERM U916, Institut Bergonié Comprehensive Cancer Centre, Univ. Bordeaux, F-33000 Bordeaux, France

² Karolinska Institutet, Radiumhemmet and Karolinska University Hospital, Stockholm, Sweden

³ Rudbeck Laboratory, Uppsala University, Uppsala, Sweden

*Correspondence to: Richard Iggo, INSERM U916, Bergonié Cancer Institute, 229 Cours de l'Argonne, 33076 Bordeaux, France. E-mail: R.Iggo@bordeaux.unicancer.fr

Abstract

We have previously tested biopsies from 1469 breast tumours with a *p53* functional assay in the context of a prospective clinical trial (EORTC 10994/BIG 1–00). The goal of the trial was to determine whether *p53* status could be used to select patients who would benefit from inclusion of taxanes in anthracycline-based chemotherapy. The results of the trial were negative. To test whether this was because the functional assay misclassified the tumours, we have reanalysed two groups of biopsies by Sanger sequencing and Roche 454 next generation sequencing (NGS). Comparison of yeast data with pooled cDNA sequencing data in an initial cohort of 69 biopsies showed that conventional sequencing is insensitive when the mutant *p53* content is low. A second cohort of 48 biopsies was used to compare directly the yeast assay with Sanger and NGS technology. The mutant sequence was difficult to detect in sequence chromatograms of pooled cDNA, whereas NGS unequivocally identified mutations in every case classified as mutant by the functional assay. The NGS data showed that small deletions, probably caused by PCR splicing, account for most of the unexplained background in the yeast assay. We conclude that mutation detection techniques that test multiple clones, such as the *p53* functional assay and NGS, are more reliable than Sanger sequencing of pooled DNA; that the high *p53* mutation rate (44%) seen with the yeast assay in the EORTC 10994/BIG 1–00 trial reflects this high sensitivity; and that NGS with Roche 454 technology could be used to identify the *p53* mutations in the remaining tumours previously tested in yeast in the EORTC10994/BIG 1–00 trial.

Copyright © 2013 Pathological Society of Great Britain and Ireland. Published by John Wiley & Sons, Ltd.

Keywords: *p53* yeast functional assay; breast cancer; next generation sequencing

Received 10 April 2013; Revised 5 July 2013; Accepted 24 July 2013

No conflicts of interest were declared.

Introduction

The *p53* gene (*TP53*) is one of the most commonly mutated genes in human tumours [1,2]. A long-standing model posits that specific *p53* mutations are linked to drug response [3]. Given the established role that *p53* plays in the response to chemotherapeutic drugs in experimental models, we organized a multi-centric phase III clinical trial in 1999 to test whether *p53* status influences the response of human breast tumours to chemotherapy. Our hypothesis was that *p53* mutant tumours would be resistant to anthracyclines but would remain sensitive to taxanes. The trial, EORTC 10994/BIG 1–00, compared a regimen containing taxanes with one without taxanes [4]. The results of the clinical trial did not confirm our hypothesis. We used a yeast functional assay to determine *p53*

status in the clinical trial. The yeast assay distinguishes inactivating mutations from biologically unimportant passenger mutations [5] but provides no information on the identity of the mutations, so we were unable to ask whether specific *p53* mutations play a role in chemotherapy resistance. To address this question, we have now performed a feasibility study to determine whether next generation sequencing (NGS) technology could be used to identify the *p53* mutations present in the cDNA that was originally used to test *p53* status in the clinical trial.

We show that NGS is more reliable than direct sequencing when the mutant *p53* mRNA fraction in the tumour is low. The reason for the high sensitivity of the yeast assay and NGS is that they test clones, which means that wild-type sequences derived from normal tissue do not obscure the mutant signal.

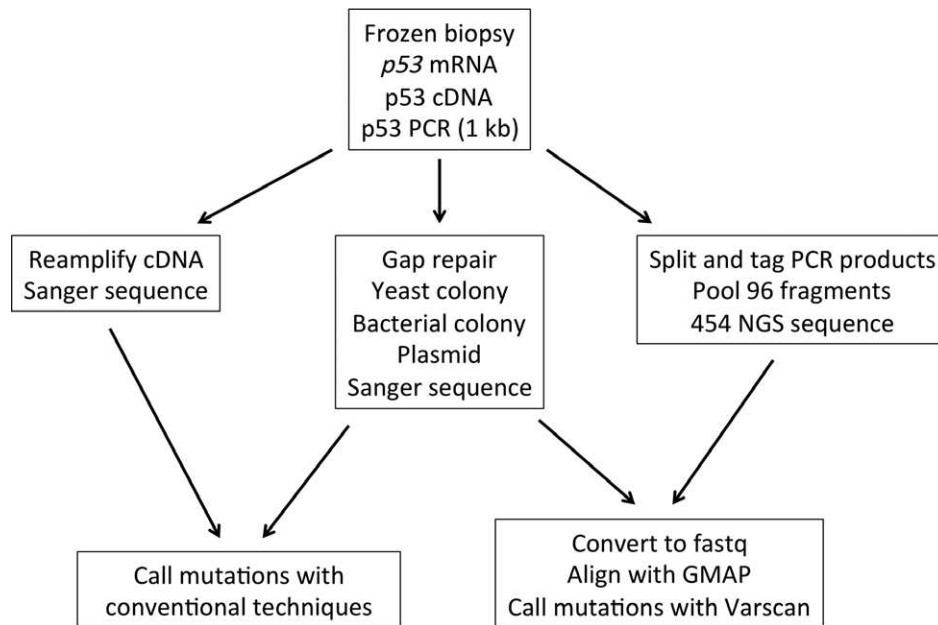


Figure 1. Pipeline for sequence analysis. Plasmids rescued from yeast were sequenced to identify the mutations present in the biopsies. The plasmid data were compared with data from direct sequencing of the PCR products or NGS from the PCR products.

Materials and methods

Consent

p53 sequencing was a secondary endpoint of the original EORTC 10994 study for which all patients gave signed informed consent as a condition for enrolment in the study [4].

DNA sequencing

The PCR products used to test *p53* status in yeast in the original study [4] were split into two fragments for NGS. Tags and 454 sequencing primers (Supplementary Table 1) were added by PCR as shown in Supplementary Figure 1 (see the Supplementary materials and methods for details). A pool containing 96 tagged PCR products was sequenced in a single batch on a GS FLX Amplicon sequencer using Titanium series chemistry (Roche 454 Life Sciences, Branford, CT, USA). Plasmids and PCR products were analysed by conventional Sanger sequencing as described in refs 4 and 6 and in the Supplementary materials and methods.

The sequencing data (fastq files) are available in the NCBI SRA database under accession number SRP020456 (BioProject PRJNA193388).

NGS analysis pipeline

The pipeline for calling variants is shown in Figure 1. The fasta and qual files were merged to give fastq files that have been deposited in the NCBI SRA database under accession number SRP020456 (BioProject PRJNA193388). The alignment and variant calling pipeline is described in the Supplementary materials and methods.

Yeast functional assay

Several mutants identified in this study were retested by yeast functional assay as previously described [5]. To test for temperature sensitivity, yeasts were incubated at 25°C, 32°C, and 36°C for 3 days.

Results

Comparison of next generation sequencing to Sanger sequencing

We initially attempted to identify *p53* mutations by Sanger sequencing of the RT-PCR products used for the yeast assay. Most mutations were missed by standard mutation detection software when the mutant fraction estimated from the yeast assay was below 40% (Supplementary Table 2). To explore the reasons for this failure, we performed a direct comparison of mutation detection techniques in a fresh cohort of 48 biopsies taken from 31 tumours. The tumours were not selected randomly or sequentially; instead, they were chosen to give a group of nine 'positive control' tumours with 59–92% red colonies in the yeast assay and a group of 12 'negative control' tumours scored as wild type in the yeast assay. To challenge the NGS technology, we selected a third group of ten 'difficult' tumours that gave 21–48% red colonies. To check for reproducibility, we tested two biopsies per tumour for 17 of the 19 tumours in the positive control and difficult groups. The mutations identified in the plasmids in the 'positive control' group were all missense or in-frame mutations, whereas the mutations in the 'difficult' group were all nonsense, frameshift or splicing mutations, apart from one in-frame deletion

Table 1. Mutations identified by NGS compared with mutations identified by Sanger sequencing

EORTC ID	Red yeast colonies biopsy 1 (%)	Red yeast colonies biopsy 2 (%)	No of plasmids with the indicated mutation	Sanger plasmid mutations	No of biopsies tested by NGS/Sanger cDNA pool sequencing	Sanger cDNA pool mutations	NGS mutations	Mutation type
169	85	77	3/3	Y163C	2	Y163C	Y163C	Missense
207	62	73	2/2	R110P	2	R110P	R110P	Missense
221	84	67	4/4	Y220C	2	Y220C	Y220C	Missense
279	79	72	4/4	C275R	2	C275R	C275R	Missense
316	92	92	4/4	F134L	2	F134L	F134L	Missense
318	67	33	nt	na	2	R282W	R282W	Missense
320	83	73	4/4	R110ΔRLGF	2	R110ΔRLGF	R110ΔRLGF	In-frame deletion
340	66	77	4/4	I195S	2	I195S	I195S	Missense
341	59	86	3/3	R273H	2	R273H	R273H	Missense
83	37	22	4/4	p53-beta	2	p53-beta*	p53-beta	Splicing
183	48	45	4/4	intron 4 sd [†]	2	Intron 4 sd [†]	Intron 4 sd [†]	Splicing
192	23	29	1/4	P64fs	2	P64fs	P64fs	Frameshift
215	38	25	2/3	S241fs	2	S241fs [†]	S241fs	Frameshift
269	40	43	3/4	Intron 6 sd	2	Intron 6 sd	Intron 6 sd	Splicing
272	36	25	2/3 [†]	Intron 6 sd	2	Intron 6 sd	Intron 6 sd	Splicing
276	27	na	2/4	G245ΔG	1	wt	G245ΔG [§]	In-frame deletion
290	22	21	2/4	S166fs	2	S166fs	S166fs	Frameshift
323	22	na	2/4	Q165Z	1	wt	Q165Z	Nonsense
326	22	23	1/4	L145fs	2	L145fs	L145fs	Frameshift
158	7	na	na	na	1	wt	wt	na
193	7	na	na	na	1	wt	wt	na
256	8	na	na	na	1	P222L	P222L	Missense (SNP)
267	8	na	na	na	1	wt	wt	na
284	5	na	na	na	1	wt	wt	na
285	13	na	na	na	1	wt	wt	na
288	10	na	na	na	1	wt	wt	na
306	6	na	na	na	1	wt	wt	na
311	10	na	na	na	1	wt	wt	na
312	7	na	na	na	1	wt	wt	na
319	10	na	na	na	1	wt	wt	na
322	8	na	na	na	1	wt	wt	na

Full details of the mutations are given in Supplementary Tables 3 and 4. [†]The plasmids contained a mixture of abnormally spliced products; the same splice variant was seen in one plasmid as in the pooled cDNA and NGS reads (the causal mutation was only visible in the plasmid sequences).

*The causal mutation was not identified.

[†]A probable PCR splicing artefact was also identified in biopsy 1.

[†]A splice donor mutation was present in one plasmid and intron retention in the other.

[§]A P191A mutation encoding functional p53 was also present in this case (see text for details).

nt = not tested; na = not applicable; wt = no mutations identified; fs = frameshift (the mutant codon is indicated); sd = splice donor.

(Table 1). This is expected because nonsense-mediated mRNA decay (NMD) reduces the level of mRNA containing premature stop codons.

NGS sequence analysis: common artefacts

The base quality, read depth, and coverage of the NGS sequences are summarized in Figure 2 and Supplementary Figure 2. As expected for 454 technology, low phred scores were mainly seen at homopolymers. The sites most commonly affected are indicated by black circles in Figure 2. We previously reported that PCR splicing artefacts can give rise to significant background mutations in the yeast assay when reverse transcription is primed with hexamers [7]. Before starting the EORTC 10994 study, we showed that use of a p53-specific reverse transcription primer substantially reduced this problem. The NGS data show that small deletions probably caused by PCR splicing still occurred to some extent in all samples in exons 4–6 (Figure 3A). To quantify the phenomenon,

we generated a 'PCR splicing score' by counting the number of reads containing deletions present in at least two tumours (Supplementary Table 1). To visualize the deletions, we plotted deletion size against position in the p53 cDNA (Figures 3B and 3C). The size of the plotted points reflects the number of reads containing the deletion. Presented in this way, it is easy to distinguish high-quality samples (Figure 3B) from those with frequent PCR splicing artefacts (Figure 3C). These artefacts were easily removed by the filters in the analysis script and did not generate false positives in the final analysis.

NGS allele depth versus % red colonies in the yeast functional assay

Figure 4 shows the fraction of mutant reads compared with the fraction of red colonies in the yeast assays. There is excellent agreement between the two measures of mutant fraction ($r = 0.97$). The y-intercept is 15.1,

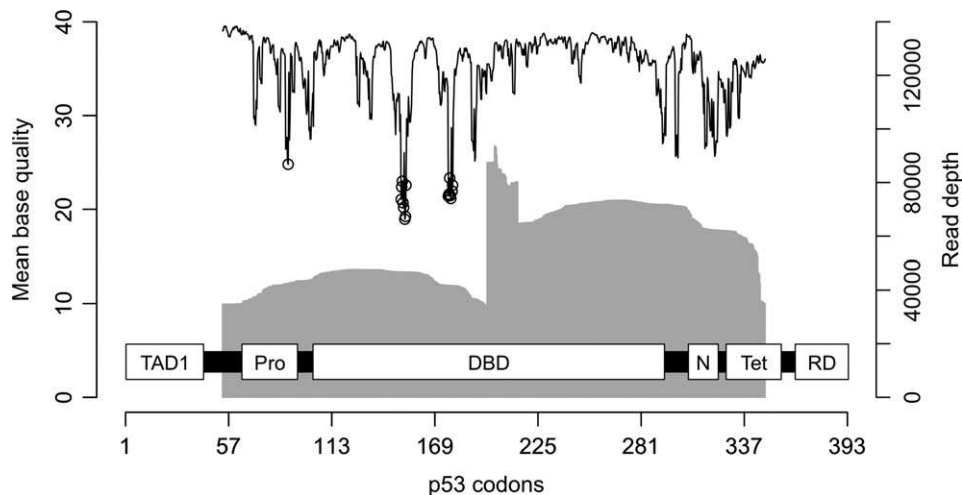


Figure 2. Sequence quality showing the effect of homopolymers in *p53*. Mean base quality (phred scale) and read depth are plotted against position in *p53*. The continuous line shows base quality. The grey silhouette shows read depth. The domain structure of *p53* is shown below the plot: TAD1, transactivation domain 1; Pro, proline-rich domain; DBD, DNA binding domain; N, nuclear localization signal; Tet, tetramerization domain; RD, regulatory domain. The region sequenced spanned codons 54–348. The black circles show 16 bases in homopolymers where the mean quality was below 25, affecting codons 89, 151, 152, 153, 176, 177, and 178.

indicating that there are no mutant reads in the NGS data when the % red colonies is 15.1%. This is slightly higher than our estimate of the background based on the % red colonies in cases scored as wild type in the published study ($11 \pm 4\%$) [4], but nevertheless remarkably similar given the fact that the current estimate is derived from NGS of a small selected population of mutant cases, whereas the previous estimate was derived from yeast assays of the full population of wild-type cases.

NGS sequence analysis: 'positive control' group

In the group of samples with high mutant RNA content, the same mutations were found in the plasmid sequences as in the NGS sequences, and the same mutations were found in both biopsies of a pair (Table 1 and Supplementary Table 3). We conclude that NGS faithfully detects abundant mutations, as expected.

NGS sequence analysis: 'negative control' group

The analysis pipeline reported no mutations in 11 of the 12 samples in the negative control group. The final sample had a P222L variant in 81% of the reads (Table 1 and Supplementary Table 3). The average base phred score was 39, indicating that the variant was called with high confidence. This variant is labelled as an SNP (rs146340390) but the dbSNP minor allele frequency is so low (0.000) that it could be a mutation. The reason for using the yeast assay was to distinguish functionally important mutations from polymorphisms and passenger mutations. To test whether the P222L mutant was partially active, we tested it in yeast at different temperatures. The colonies were white at 25–36°C, but after storage at 4°C they gradually became slightly pink (Figure 5), indicating that the P222L mutant protein has a mild transcriptional defect.

We conclude that NGS identifies wild-type samples with high confidence but additional filtering is required to distinguish functionally important mutations from polymorphisms and passenger mutations.

NGS sequence analysis: 'difficult' group

The 'difficult' tumours included four with nonsense, frameshift or in-frame deletion mutations and two for which we could not confidently identify the mutation in the plasmid sequences. We identified nonsense or frameshift mutations by NGS in all ten biopsies from these tumours (Table 1 and Supplementary Table 3). Nine of these biopsies had yeast results in the range 21–29% red colonies, close to the 20% cut-off used to score tumours as mutant. NGS thus demonstrates that tumours just above the cut-off in the yeast study were correctly scored as mutant. The remaining four tumours in the 'difficult' group had splicing mutations. These mutations commonly lead to a mixture of intron retention and exon skipping; the actual mutation is typically only visible in the former. This is exactly what was seen in the NGS data. The mutant base was present in the plasmid sequences from all four tumours but only present in the NGS data in three of them. In the fourth tumour (183, Supplementary Table 3), the mutation was located five nucleotides after the splice site (cg[^]gtcag>cg[^]gtcac). It was visible in the three plasmids with intron retention. In contrast, it was not present in the NGS sequences because none contained the retained intron, which increases the size of the fragment by 757 bp. Instead, there were multiple reads using a cryptic donor site in exon 4; this variant was also seen in one plasmid. The cryptic donor had the sequence ag[^]gt. Since this corresponds to a normal splice donor, the alignment software labelled it as a normal intron. To solve this problem, we aligned the NGS sequences to specific *p53* mRNA isoforms

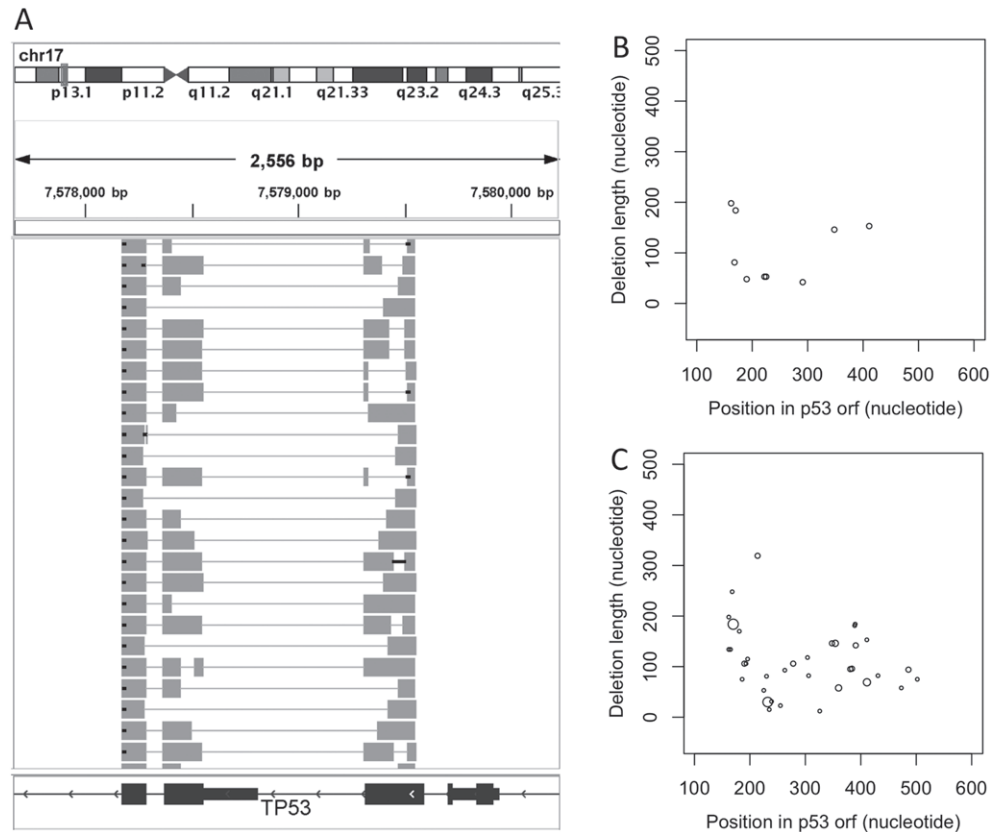


Figure 3. Artefactual deletions caused by PCR splicing. PCR splicing creates artefactual deletions in exons 4–6. (A) Reads from a sample classified as wild type (158) displayed in the Integrated Genome Viewer (IGV). The thick bars in the bottom panel show *p53* coding exons. The light grey bars in the middle panel show sequencing reads. Gaps in reads that do not begin at exon boundaries are probably caused by PCR splicing. (B, C) Plots showing the position and length of deletions in the biopsies from tumour 340. Biopsy 2 has few deletions (B), whereas biopsy 1 has frequent deletions (C). The size of the dots is proportional to the number of mutant reads. Deletions present in at least two tumours were used to generate a PCR splicing score (see Supplementary Table 1).

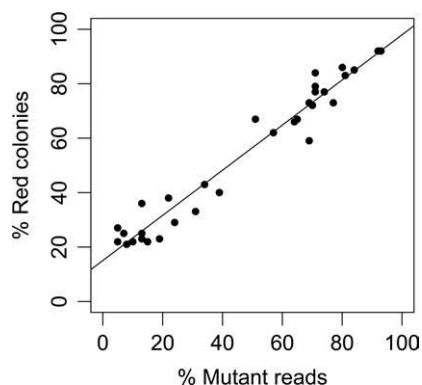


Figure 4. Estimating the background in the yeast assay. The percentage of NGS reads containing the mutations in Table 1 is plotted against the percentage of red yeast colonies. Reads containing P191A in sample 276 and P222L in sample 256 were excluded because these variants do not give red colonies. Cases 83 and 183 were excluded because mutant reads containing splice variants that increased the length of the template were lost from the NGS data. The NGS data are derived from the alignment to the prototypic *p53 α* mRNA.

and forced the software to report unaligned bases as deletions. The mutation was efficiently called with this strategy (Table 1 and Supplementary Table 3). The alternate splice in case 83 is caused by a mutation in

the donor site at the end of the alternative exon that improves the donor site ($tt^{\wedge}gt > tg^{\wedge}gt$), leading to more frequent inclusion of the alternative exon. All four plasmids contained the mutation but it was only called in one of the biopsies in the NGS data. The obstacle to calling the mutation in the other biopsy was the low allele depth (two reads with the beta and two with the gamma isoform). The splicing mutations in the two remaining cases were correctly identified (Table 1 and Supplementary Table 3). In both cases, the mutation affected the intron 6 splice donor site, resulting in the use of a cryptic donor ($tg^{\wedge}gt$) five nucleotides after the normal donor. The pathologically spliced forms were not called as such after alignment to the genomic sequence, although the mutation itself was called, but they were correctly identified in the alignment to mRNA. In summary, all of the mutations in the ‘difficult’ group were correctly identified by NGS.

Direct comparison of NGS with Sanger sequencing

To allow direct comparison of NGS with Sanger sequencing, we performed Sanger sequencing of the pooled cDNA PCR products in the second cohort and asked an accredited diagnostic lab geneticist to identify the mutations by visual inspection of the

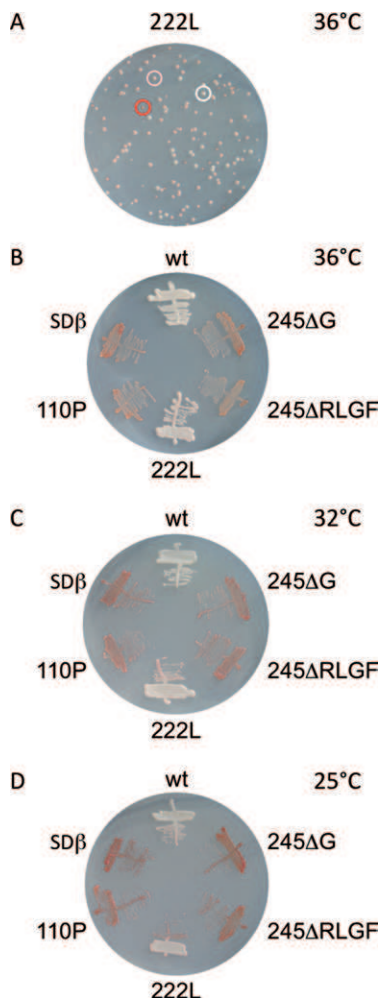


Figure 5. Yeast functional assay for the P222L p53 mutant. The yeast reporter strain contains an *ADE2* gene expressed from a *p53*-responsive promoter. (A) Yeast transfected with the *p53* PCR product from case 256. The pink circle shows a pink colony containing the P222L mutant; the white circle shows a white colony containing wild-type *p53* derived from normal tissue; the red circle shows a red colony containing mutant *p53* inactivated by a PCR mutation. (B–D) Yeasts expressing some of the mutants identified in this study were grown at the temperatures indicated. The P222L mutant colonies are slightly cold-sensitive. The R110P mutation is labelled as an SNP (rs11540654) but is clearly inactive.

chromatograms. Despite the low mutant content in some samples, all but two of the mutations were identified correctly (Table 1 and Supplementary Table 4). The two mutations missed were present in the NGS data in 5% of the reads. Figure 6 shows three examples of chromatograms with low mutant content in which the mutation was nevertheless correctly identified. The phred scores calculated with Tracemaker software are shown beneath the mutant peaks. They indicate that the probability of error in automated calling of the mutant bases was commonly in the range 10–30%. This should be compared with the phred = 30 cut-off for calling mutant peaks in the NGS data, which corresponds to a probability of error of 0.1%. We conclude that NGS allows much more confident identification of mutations than Sanger sequencing in samples heavily contaminated with normal tissue.

Discussion

The main conclusion from this study is that mutation detection techniques that test multiple clones, such as the yeast assay or NGS, are more sensitive than direct sequencing of bulk tumour DNA. The NGS data also indicate that the tumours in the EORTC 10994 clinical trial scored as mutant by the yeast assay probably do contain authentic *p53* mutations. The mutation rate based on the yeast assay was 44%, a higher figure than was anticipated by the study design [4]. Analysis of the distribution of red colonies showed two mutant peaks, one at 76% and the other at 29% [4]. The latter was uncomfortably close to the cut-off for scoring tumours as mutant (20% red colonies). NGS shares with the yeast assay the fact that it tests multiple clones, but has the advantage that background mutant clones (reads) can be distinguished from one another by their sequence. This means that the cut-off can be pushed much lower in studies using NGS than in studies using the yeast assay. The NGS results indicate that samples near the threshold in the yeast assay contained genuine mutations. Taken together, the results indicate that NGS is more reliable than Sanger sequencing of pooled cDNA when the mutant RNA content is low, as is often the case in clinical samples. We suspect that many studies in the older literature may have underestimated the *p53* mutation rate for this reason.

In the conditions used for the clinical trial, we expected PCR mutations to generate a background of about 3.5% red colonies [8] but the observed background was $11\% \pm 4\%$ [4]. The NGS data show us the likely source of much of the unexplained background. The PCR analogue of DNA repair by microhomology-mediated end joining is called PCR splicing. It occurs when template breaks terminate replication before the polymerase reaches the opposing primer site. The termini of the resulting fragments anneal to one another in ensuing cycles at sequences with a few bases of homology. We previously reported that this was a major source of background in the yeast assay when reverse transcription was primed with hexamers [7]. It is substantially reduced by priming reverse transcription with a *p53*-specific primer, but the NGS data show that it still occurred. The diversity of the variants generated by PCR splicing means that they were almost all removed automatically by the filters in Varscan, but to make the pipeline more robust we used a Perl script to explicitly remove the commonest PCR splice variants. The Varscan filters in the script are stringent for base quality (phred ≥ 30) but relatively permissive for mutant read depth (≥ 5 mutant reads, $\geq 4.5\%$ mutant reads) because we were interested in identifying mutants that scored near the background in the yeast assay. At these thresholds, there was one case (276) in which the script returned two mutations. The second mutation (P191A) was present in 11% of the NGS reads, but is probably a passenger mutation, a PCR mutation, or a sequencing artefact. The mutant

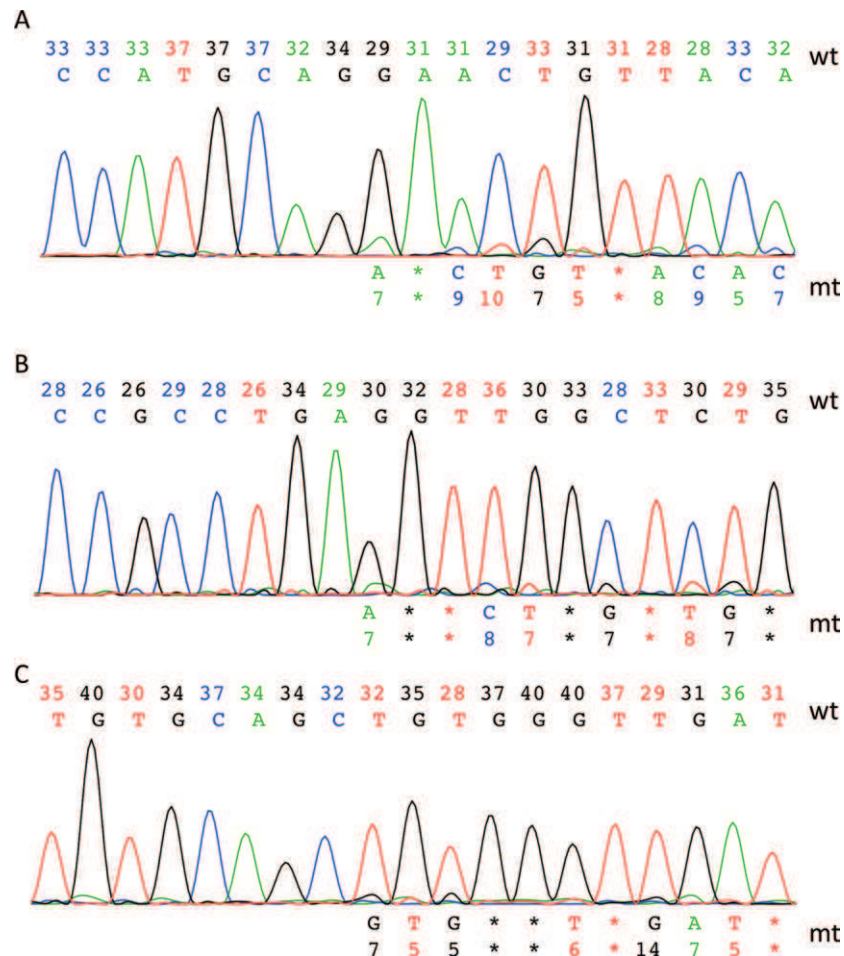


Figure 6. Sequence chromatograms showing secondary peaks. Sanger sequencing was performed on pooled cDNA from biopsies 215.2 (A), 272.2 (B), and 326.2 (C). The mutant content of these biopsies estimated from the NGS data was 14%, 7%, and 13%, respectively. The base calls and quality (phred) scores are shown for the wild-type and mutant sequences above and below the chromatograms, respectively. An asterisk in the mutant track indicates that the base was identical in both sequences.

protein is wild type in the Ishioka functional assay database [9], and the mutation is adjacent to a deletion with a low phred score, indicating that the NGS data are potentially unreliable at this base.

One argument for using the yeast assay in the clinical trial was that it would distinguish functionally important mutations from passenger mutations and polymorphisms. Interestingly, the negative control group in the present study contained a mutation with exactly those properties. To exclude the possibility that the sample was wrongly scored, we retested it in yeast and found it to have a very mild transcriptional defect. According to a p53 function predictor [10], the P222L mutation should have 24% activity, a value that would place it in the mutant camp, but the Ishioka database shows it to be active on most binding sites [9], consistent with our results (Figure 5).

We chose to sequence the p53 cDNA PCR products used for the clinical trial because genomic DNA was not available. Alignment to genomic DNA with GMAP worked perfectly but the variant calling software was unable to identify pathological splice variants in the GMAP output because the cryptic donor sites were similar or identical to normal splice sites, so the

corresponding intron bases were set by GMAP to N in the CIGAR string. The variants were definitely abnormal because we could see the causal mutations in the plasmids and they do not correspond to any of the rare variants reported in a high-depth RNA-seq study on p53 [11]. To identify them, we aligned the NGS reads to the p53 mRNA and forced GMAP to report unaligned bases as deletions. A more pernicious problem was the loss of fragments containing retained introns or additional exons. This reduced the mutant allele depth for one mutant (the p53 beta splice donor mutation in case 83) to the range that overlaps with trivial NGS artefacts. It remains to be seen whether this problem is tractable because allele depth is such a useful criterion for filtering NGS data.

Our ultimate goal is to sequence p53 from the 1469 tumours tested in yeast in the EORTC 10994 trial, to find out whether specific p53 mutations predict the response of tumours to chemotherapy. We chose to test 454 NGS technology because we could incorporate tags for 96-fold multiplexing in a simple PCR step starting from our existing PCR products. While we were able to identify all of the mutations that we sought, few scientists could view the profile in Figure 2

with total equanimity. In addition to its inability to sequence homopolymers, 454 technology has a rather low capacity by current standards. Were the read length of competing technologies to reach ~550 bp, libraries based on our existing strategy could easily be further multiplexed to sequence all 1469 tumours in one run.

The three main conclusions from this study are that it would be possible to identify the *p53* mutations in the tumours from the EORTC 10994 clinical trial with the protocol described here; that the tumours from the trial that were scored as mutant in the yeast assays [4] probably do contain authentic *p53* mutations; and that studies from the pre-NGS era probably underestimate *p53* mutation rates.

Acknowledgments

We thank Monica Wallin for technical assistance; the bioinformaticians who generously made available their software through public websites; the members of the Institut Bergonie breast cancer group for helpful advice; and Macha Nikolski for critical reading of the manuscript. We also thank the French Cancer League ('Equipe Labellisée Ligue Contre le Cancer 2011'), the French National Research Agency (ANR-08-CEXC-016-01), and the Fondation Widmer pour la Lutte Contre le Cancer for financial support.

Author contribution statement

The authors contributed in the following way: TS, NS, and EM: Sanger sequencing; EM: preparation of libraries; TS, JR, and RI: data analysis; HB, JB, and RI: study design; RI: yeast assays and writing of the manuscript. All of the authors approved the final version of the manuscript.

References

- Petitjean A, Mathe E, Kato S, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* 2007; **28**: 622–629.
- Goh AM, Coffill CR, Lane DP. The role of mutant p53 in human cancer. *J Pathol* 2011; **223**: 116–126.
- Aas T, Borresen AL, Geisler S, et al. Specific P53 mutations are associated with *de novo* resistance to doxorubicin in breast cancer patients. *Nature Med* 1996; **2**: 811–814.
- Bonnefoi H, Piccart M, Bogaerts J, 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* 2011; **12**: 527–539.
- Flaman JM, Frebourg T, Moreau V, et al. A simple p53 functional assay for screening cell lines, blood, and tumors. *Proc Natl Acad Sci U S A* 1995; **92**: 3963–3967.
- Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006; **314**: 268–274.
- Waridel F, Estreicher A, Bron L, et al. Field cancerisation and polyclonal p53 mutation in the upper aero-digestive tract. *Oncogene* 1997; **14**: 163–169.
- Flaman JM, Frebourg T, Moreau V, et al. A rapid PCR fidelity assay. *Nucleic Acids Res* 1994; **22**: 3259–3260.
- Kato S, Han SY, Liu W, et al. Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A* 2003; **100**: 8424–8429.
- Soussi T, Hamroun D, Hjortsberg L, et al. MUT-TP53 2.0: a novel versatile matrix for statistical analysis of TP53 mutations in human cancer. *Hum Mutat* 2010; **31**: 1020–1025.
- Mercer TR, Gerhardt DJ, Dinger ME, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotechnol* 2012; **30**: 99–104.
- Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnol* 2011; **29**: 24–26.
- Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 2005; **21**: 1859–1875.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article.

Supplementary materials and methods.

Table S1. Primers used for multiplexing showing the PCR splicing score for each biopsy and the number of reads per tag.

Table S2. Mutations identified by Sanger sequencing in cohort 1.

Table S3. Mutations identified by NGS in cohort 2.

Table S4. Mutations identified by Sanger sequencing in cohort 2.

Figure S1. Multiplex PCR strategy.

Figure S2. NGS read quality before and after trimming.

Supplementary Data 1. Bam files for NGS and plasmid sequences. The fastq files containing the raw data are available in the NCBI SRA database under accession number SRP020456 (BioProject PRJNA193388).