TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance

Laxmi Silwal-Pandit¹,²*, Hans Kristian Moen Vollan¹,²,³*, Suet-Feung Chin⁴,⁵, Oscar M. Rueda⁴,⁵, Steven McKinney⁸,⁹, Tomo Osako⁸,⁹, David A. Quigley¹,²,⁶, Vessela N. Kristensen¹,²,⁷, Samuel Aparicio⁸,⁹, Anne-Lise Borresen-Dale¹,²,⁴**, Carlos Caldas⁴,⁵,¹⁰,¹¹**, and Anita Langerød¹,²**

*Shared first authors
**Shared corresponding authors

¹Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Norway
²The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Norway
³Department of Oncology, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Norway
⁴Cancer Research UK, Cambridge Institute, UK
⁵Department of Oncology, University of Cambridge, UK
⁶Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, USA
⁷Department of Clinical Molecular Oncology, Division of Medicine, Akershus University Hospital, Norway
⁸Department of Pathology and Laboratory Medicine, University of British Colombia, Vancouver, Canada
⁹Molecular Oncology, British Colombia Cancer Research Center, Vancouver, Canada
¹⁰Cambridge Experimental Cancer Medicine Centre, UK
¹¹Cambridge Breast Unit, Addenbrooke’s Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, UK

Corresponding authors
Anita Langerød (anita.langerod@rr-research.no), Anne-Lise Borresen-Dale (a.l.borresen-dale@medisin.uio.no)
and Carlos Caldas (Carlos.Caldas@cruk.cam.ac.uk)

Running title
TP53 in breast cancer

Keywords
TP53, mutation, LOH, breast cancer, molecular subtypes

Grant Support
This work was supported by, Norwegian Cancer Society (0332), South-Eastern Norway Regional Health Authority (2011079), The Research Council of Norway (163027/V40), The Norwegian Radium Hospital Foundation, EuroCan Platform FP7 (260791), The British Columbia Cancer Foundation, and Cancer Research UK.

Disclosure of potential conflicts of interest
No potential conflicts of interest are disclosed

Word count:
4389
Statement of translational relevance

Somatic mutation in the TP53 gene is a strong prognostic marker in breast cancer, but the clinical and biological impact in subtypes has not been clear due to small patient cohorts, suboptimal methods to assess mutation status and lack of sub-typing by molecular profiling. The current study presents analyses of somatic TP53 mutations in 1420 breast cancer patients from the METABRIC cohort(1), and is the largest to date exploring the effect of TP53 mutations in a subtype specific manner. The study shows that the TP53 mutational spectrum and the prognostic implications indeed are subtype specific. This knowledge is crucial with respect to the potential clinical application of TP53 as a prognostic and predictive marker, for optimal design of clinical trials, and in further development of TP53 targeted therapies.
Abstract

**Purpose:** In breast cancer, the *TP53* gene is frequently mutated and the mutations have been associated with poor prognosis. The prognostic impact of the different types of *TP53* mutations across the different molecular subtypes is still poorly understood. Here, we characterize the spectrum and prognostic significance of *TP53* mutations with respect to the PAM50 subtypes and Integrative Clusters (IC). **Experimental design:** *TP53* mutation status was obtained for 1420 tumor samples from the METABRIC cohort by sequencing all coding exons using the Sanger method. **Results:** *TP53* mutations were found in 28.3% of the tumors, conferring a worse overall and breast cancer specific survival (HR=2.03, 95%CI=1.65-2.48, p<0.001), and were also found to be an independent marker of poor prognosis in estrogen receptor positive cases (HR=1.86, 95%CI=1.39-2.49, p<0.001). The mutation spectrum of *TP53* varied between the breast cancer subtypes, and individual alterations showed subtype specific association. *TP53* mutations were associated with increased mortality in patients with Luminal B, HER2-enriched and Normal-like tumors, but not in patients with Luminal A and Basal-like tumors. Similar observations were made in ICs, where mutation associated with poorer outcome in IC1, IC4 and IC5. The combined effect of *TP53* mutation, *TP53* LOH and *MDM2* amplification on mortality was additive. **Conclusion:** This study reveals that *TP53* mutations have different clinical relevance in molecular subtypes of breast cancer, and suggests diverse roles for TP53 in the biology underlying breast cancer development.
Introduction

Three decades of research on TP53 have documented its fundamental role as a regulator of key cellular processes involved in controlling proliferation and in maintaining the integrity and stability of the genome(2-4). The TP53 tumor suppressor protein is activated in response to a variety of stress signals and suppresses cellular transformation by triggering cell cycle arrest, DNA repair and apoptosis. In addition, a role for TP53 in processes such as metabolism, fertility, angiogenesis, immune responses and stem cell maintenance has been shown(4).

The importance of TP53 in tumor progression is evidenced by the high mutation frequency found in many cancer types, including breast cancer. Point mutations are the most common somatic aberration, followed by small insertions and deletions. The mutations are mostly missense and are predominantly located in exons 5-8, spanning the DNA binding domain of the protein(5). Mutations may cause complete or partial loss of protein function, acquisition of dominant negative effect, or gain of function(6,7). In addition to mutations, the TP53 pathway may also be disrupted by allelic deletions (loss of heterozygosity; LOH) of TP53, or amplification of the TP53 regulator MDM2.

Based on differential gene expression, breast cancer can be classified into five subgroups (Luminal A, Luminal B, Basal-like, HER2-enriched and Normal-like) with distinct biology and clinical outcome(8,9). These subtypes (PAM50) have been reproduced in several independent studies(10-12). An important development in clinical implementation is the recent recognition of the relevance of subtypes, as implemented in the St. Gallen International Breast Cancer Guidelines, and the development of single sample predictors to classify individual tumors (Nanostring® Technologies)(13). As clinical outcome even within subtypes
is variable, a more refined taxonomy of breast cancer is still needed. Recently, copy number and gene expression data from 2000 breast cancer patients from the METABRIC (Molecular taxonomy of breast cancer international consortium) cohort, was published(1). Integrative clustering of the top 1000 cis-driven genes identified ten subtypes (Integrative Clusters; IC) with distinct genomic drivers and clinical outcome(14).

Around 30% of breast cancers have been reported to harbor a somatic mutation in the TP53 gene. TP53 mutation status is a strong marker of prognosis, while its predictive value is debated(15,16). Small clinical cohorts, suboptimal methods used to assess TP53 mutations status and lack of consistency in how the mutations are classified, have hindered a robust correlation with clinical features(15-18). The frequency and type of mutations varies across the PAM50 subtypes, but the clinical impact is unclear(11,19,20). Investigating the association of TP53 mutations both with PAM50 and the recently described ICs is therefore important.

Here, we report somatic TP53 mutations in 1420 breast cancer patients from the METABRIC cohort and show that both mutational spectrum and prognostic associations are subtype specific. The differences are discussed in relation to important biological features, such as TP53 LOH, MDM2 amplification, genomic instability, mutational process and evolutionary selection in the tumor. The overall aim of the study was to create a more solid fundament for clinical trial design and for future implementation of TP53 as a biomarker.
Material and Methods

Patient material

TP53 mutation status was successfully obtained for 1420 tumor samples from the METABRIC cohort(1). Detailed inclusion/exclusion criteria are shown in the patient flow diagram (Fig. S1). The fresh frozen specimens from primary invasive breast carcinomas prior to adjuvant treatment were collected from five different sites in the UK and Canada after ethical committee approval (Addenbrooke’s Hospital, Cambridge; Guy’s Hospital, London; Nottingham; Vancouver; Manitoba), along with clinical and pathological annotation. The mean age was 60.5 at time of diagnosis/surgery (range 21.9-96.3 years) and mean follow-up was 7.6 years (range 0.1-24.5 years). The Kaplan-Meier plots were truncated at 15 years.

Sequencing analysis

Extraction of nucleic acids was performed as previously described(1). TP53 mutation status was assessed by sequencing the entire coding region (exons 2-11), including splice junctions. The sequencing was performed according to the manufacturer’s procedures using BigDye® Terminator v1.1 Cycle Sequencing kit (Life Technologies™), which applies the Sanger sequencing chemistry. The samples were run on the 3730 DNA Analyzer (Life Technologies™), a capillary electrophoresis based automated DNA sequencer.

The sequences had been previously analyzed automatically, which underestimated mutation calls(1). Therefore here the electropherograms were inspected manually, assisted by the SeqScape v2.5 software (Life Technologies), by two experienced operators. Any discrepancies in mutation calls were resolved by re-sequencing the samples.

The probability of occurrence of a mutation was estimated for each exon separately in samples with all exons successfully sequenced. Samples with one or more exons
unsuccessfully sequenced and no mutations detected were excluded from analysis if the probability of having a mutation was >5% (n=22). Patients with double mutations (n=7) were excluded from the statistical analyses due to the uncertain contribution of each mutation. These 29 samples were not included in the final 1420 dataset. Survival analysis included 1404 patients, due to patients lost to follow up (n=16).

Mutation classification

The mutations were classified according to predicted effect on the protein(16); missense in DNA-binding motif (DBM), missense outside DBM, and non-missense mutations (including splice, inframe, frameshift and nonsense mutations).

Molecular sub-classification and genomic instability scoring

Subtype assignment (PAM50 and Integrative Clusters) was done as described previously(1). Copy number (Affymetrix SNP 6.0) and expression data (Illumina HT 12_v3) are deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number EGAS00000000083.

Scoring of genomic instability from SNP6 data was performed using different approaches. The Complex Arm Aberration Index (CAAI) captures short genomic regions with highly complex rearrangements such as firestorms (21). Genomic Instability Index (GII) is the proportion of amplified or deleted genomic loci(22). TP53 and APOBEC3B expression levels were measured using the ILMN_1779356 and ILMN_2219466 probes, respectively. Allele-specific copy number analysis of tumors (ASCAT), which corrects for tumor percentage and ploidy, was used to score allele specific copy number of TP53 and MDM2(23).
Lymphocyte infiltration and TP53 immunohistochemistry scoring

The degree of lymphocytic infiltration (LI) was assessed in samples with adequate morphology on H&E stained fresh frozen tissue sections (1105 of 1420 samples). The tumors were classified as “absent” (no infiltration), “mild” (light scattering of lymphoid cells), or “severe” (dense infiltration of lymphoid cells forming confluent sheets).

TP53 immunostaining was performed using a monoclonal antibody (clone DO-7, dilution 1:100; Dako, Glostrup, Denmark) on 4-μm-thick FFPE tissue microarray (TMA) sections. The labeling index value was evaluated by discretizing the percentage (1-100%) of positive nuclei within the invasive portion of each core of the TMA. A total of 869 cases had intact cores with sufficient tumor cells for TP53 scoring.

Statistical analysis

All statistical analysis were performed using R version 2.15(24) with the package ‘rms’(25) and SPSS 18.0 software (SPSS Inc., Chicago). Pearson’s chi-square, Fisher’s exact test, t-test or Kruskal-Wallis test were used when appropriate to test association between different variables. Breast cancer specific survival (BCSS), defined as the time of diagnosis to the time of a breast cancer-related death; and overall survival (OS), defined as the time of diagnosis to the time of death from any cause, were used for the survival analyses. The Kaplan-Meier estimator with log-rank test for significance, and Cox proportional hazard (PH) models, were used for the survival analysis. Inspection of standard log-log plots and tests based on correlation between time and Schoenfeld’s partial residuals were used to evaluate violations of the assumption of PH in the multivariate models(26).
**Results**

**Characterization of TP53 mutations**

TP53 was mutated in 28.3% (n=402) of 1420 cases (Table S1). The majority of the mutations were single base substitutions (n=295, 73.4%), followed by small deletions (n=75, 18.7%) and insertions (n=21, 5.2%). Complex mutations, comprising both deletions and insertions, (n=8, 2.0%) and tandem mutations (n=3, 0.7%) were uncommon. Eight (2.7%) of the base substitutions were silent/synonymous and were considered wildtype in further analyses, giving a total mutation frequency of 27.7% (n=394). The single base substitutions (n=287; silent excluded) were predominantly G:C>A:T transitions (n=142, 49.5%), frequently at CpG sites (88/142, 62%), and A:T>T:A transversions were least common (n=13, 4.5%). The proportion of G:C>A:T at CpG sites was highest in tumors with Basal-like phenotype (Fig. S2).

The distribution of the mutations was non-uniform across the gene with 81% of the mutations clustering in exons 5-8, mostly spanning the DNA binding domain of the protein. Exons 4 and 10 also harbored a substantial number of mutations, with 9.6% and 6.5% of the total count respectively. Only 2% of the mutations were in exon 9, and the remaining 1% were in exons 2, 3 and 11 (Fig. 1 and Table S1). Alterations at putative splice sites (defined here as 2 nucleotides before and after an exon) were detected in 17 cases, and the majority resided between exons 5 and 8. Mutational hotspots(5) were observed in codons 175, 179, 196, 213, 245, 248, 273, 278, 285 and 306. The mutations were mostly G:C>A:T transitions, which in codons 196, 213 and 306 resulted in nonsense codons due to the specific sequence context.

Interestingly, frameshift mutations were evenly distributed along the gene with no hotspots, whereas missense, nonsense (both with several hotspots), inframe and splice mutations were
mainly located in the DNA binding domain (Fig. 1 and Table S1). Several types of mutations, including a cluster of missense mutations, hit the oligomerization domain; these potentially prevent the formation of tetramers. Nonsense and missense mutations were nearly mutually exclusive with respect to hotspot mutational sites. This observation cannot solely be explained by codon context. The high mutability of C to T, likely followed by selection may explain this pattern.

A systematic overview of the consequences of different types of TP53 mutations on both mRNA and protein expression in a large cohort of breast cancer is currently not available. Hence, we stratified the tumors by mutation status, and found a high correlation between TP53 mRNA levels and protein expression, but only in mutated tumors (p<0.001, Kruskal-Wallis test; Fig. S3-B). Tumors with missense mutations were predominantly IHC positive and had higher mRNA levels, whereas frameshift and nonsense mutations were predominantly IHC negative with lower mRNA levels (Fig. S3-A and S3-B). Half of the wildtype tumors also showed positive IHC, but the staining intensity was relatively weak and not correlated with mRNA levels. These results show that IHC is an inadequate surrogate method for mutation screening.

Spectrum of TP53 mutations in molecular subtypes

TP53 mutations were differentially distributed in PAM50 subtypes (p<0.001, \( \chi^2 \)-test, Table 1), which confirms previous reports(11,19). The subtypes with high mutation frequency, Basal-like (65%) and HER2-enriched (53%), showed enrichment of non-missense mutations, whereas the Luminal B type showed a high proportion of missense mutations, of which the majority affected the DBM (p=0.061, \( \chi^2 \)-test) (Fig. 2A).
A notable difference of mutational hotspots was observed between subtypes. The Basal-like subtype was enriched for multiple hotspots, while Luminal A tumors had a flat mutation profile. The most frequently mutated codon, 248, was identified as a hotspot in Luminal B, HER2-enriched and Basal-like tumors, while mutations in codon 175 and 273 were observed mostly in the Basal-like subgroup. Interestingly, the nonsense mutation R213* was a hotspot in Basal-like (seven out of nine samples; Fig. 2A and 2B). This intriguing finding is also observed in other cohorts, where 16 of 20 samples with the R213* mutation are Basal-like (data not shown).

The distribution of TP53 mutations in the 10 Integrative Clusters (ICs) also revealed significant variation in frequency (p<0.001, χ²-test; Table 1). IC10 had the highest mutation frequency (76.5%), and the mutational spectrum observed in Basal-like tumors was recapitulated, suggesting association with the core Basal-like subset(14). The proportion of IC9 tumors (48%) with TP53 mutations is noteworthy since this cluster primarily consists of ER positive/HER2 negative cases. The distribution of TP53 mutations across the 10 clusters was validated in the TCGA dataset by performing analysis of deviance. This confirmed IC9, an ER positive/HER2 negative subtype, as frequently mutated (Fig. S4).

We observed that high APOBEC3B expression level was associated with TP53 mutations (p<0.001, t-test). Interestingly, this effect was strongest in the Basal-like tumors (Fig. S5), suggesting a subtype related mechanism contributing to TP53 mutagenesis. The DNA cytosine deaminase APOBEC3B has recently been suggested as a driver of mutations, especially C-to-T and C-to-G substitutions at TC motifs with non-methylated cytosines, and was shown to be over-expressed in TP53 mutated cell-lines and tumors(27).
Prognostic implications of *TP53* mutations in breast cancer

Breast cancer patients with a somatic *TP53* mutation had significantly inferior breast cancer specific survival (BCSS), [Hazard Ratio (HR) = 2.03, 95% Confidence Interval (CI) = 1.65-2.48, p<0.001, Cox Regression model, Fig. 3A, Table S2] and overall survival (OS, HR=1.59, 95%CI=1.34-1.89, p<0.001, Fig. S6-A). *TP53* mutations were associated with increased mortality also in the subset of patients that did not receive cytotoxic treatment (n=1078, HR=2.20, 95%CI=1.69-2.86, p<0.001, Fig. S6-B), which suggests that the effect is not related to response of adjuvant chemotherapy. Importantly, stratification of cases showed that the prognostic effect of *TP53* was limited to ER positive disease (Fig. 3C and 3D). As ER positive and negative breast cancers are widely recognized as different biological entities, separate multivariate survival analyses were performed. After correcting for other covariates, *TP53* mutation status was an independent predictor of outcome in ER positive patients only, both for BCSS (HR=1.86, 95%CI=1.39-2.49, p<0.001, Table S3) and for OS (HR=1.49, 95%CI=1.17-1.90, p<0.001, Table S4).

The large number of cases allowed survival analysis stratified by molecular subtypes. *TP53* mutations were associated with a worse outcome in Luminal B (HR=1.66, 95%CI=1.14-2.42, p=0.007), HER2-enriched (HR=1.69, 95%CI=1.04-2.73, p=0.032) and Normal-like subtypes (HR=3.62, 95%CI=1.67-7.88, p=0.001), whereas no significant effect was observed in the Basal-like and Luminal A subtypes (Fig. 4). *TP53* mutations conferred higher mortality in IC1 (HR=2.00, 95%CI=1.01-3.95, p=0.045), IC4 (HR=2.08, 95%CI=1.19-3.63, p=0.009) and IC5 (HR=1.88, 95%CI=1.13-3.14, p=0.015), although the number of events limits the statistical robustness of these results (Fig. S7). IC10 corresponds mostly to triple negative cases from the Basal-like subtype(1,14), and no prognostic effect was observed in this highly *TP53*
mutated (76.5%) group. Taken together these novel observations indicate that the clinical significance of mutant TP53 varies across molecular subtypes of breast cancer.

Survival differences could not be detected in patients harboring different types of TP53 mutations (Fig. 3B). Although patients with tumors bearing missense mutations in the DBM tended to have inferior outcome, this was not statistically significant.

TP53 and genomic instability

TP53 mutation is associated with genomic instability (28). TP53 mutated tumors had significantly higher rate of genomic instability (GII) (p<0.001, Kruskal-Wallis test), most distinct in Basal-like cases. Focal complex alterations are believed to be associated with telomeric attrition and a result of breakage-fusion-bridge cycles, and suggest a different mechanism for instability of the genome (21,29). Higher rates of these complex alterations (‘firestorms’), measured by CAAI, were observed in mutated tumors (p<0.001, χ²-test); predominantly in the Basal-like and HER2-enriched subgroups).

We further investigated whether tumors acquire genomic instability through alternative mechanisms of TP53 pathway deregulation (TP53 LOH and MDM2 amplification).

Significantly higher frequencies of TP53 LOH and MDM2 amplification were observed in mutated vs. wildtype tumors (p<0.001, χ²-test; Table 1), with no clear association with the type of mutation. LOH was observed in 80.8% of the mutated tumors, independently of molecular subtype. However, in TP53 wildtype cases the frequency of LOH varied across subtypes (p<0.001, χ²-test); 52% in Luminal B vs. 24% in Basal-like (Fig. S8-A). The difference was even more prominent across ICs; 80% in IC1 compared to 35% in IC10 (Fig. S8-A). These alternative mechanisms of TP53 pathway deregulation increased the mortality
in patients with wildtype tumors (LOH vs. no LOH: HR=1.56, 95%CI=1.20-2.02, p<0.001; MDM2 amplification vs. no MDM2 amplification: HR=1.86, 95%CI=1.43-2.42, p<0.001), but not in patients with mutated TP53 (Fig. S8-B). Interestingly, the combined influence of these events (TP53 mutation, TP53 LOH, or MDM2 amplification) conferred increasingly unfavorable prognosis (HR=1.54, 95%CI=1.39-1.70, p<0.001; Fig 5A) and tumors showed correspondingly higher genomic instability (p<0.001, Kruskal-Wallis test, Fig 5B).
Discussion

The current study presents analyses of somatic TP53 mutations in a large cohort of 1420 breast cancer cases with extensive clinical and molecular annotation. Our previous study was of significant size (n=1794), but a large proportion of the samples (64%) were only screened for mutations in exons 5-8 and molecular subtype information was not available(16). The TCGA dataset includes 826 samples with TP53 mutation status, but survival data is extremely limited(10). This makes METABRIC the largest cohort exploring the effect of TP53 mutations in breast cancer in a subtype specific manner. The large sample size allowed FDR correction (data not presented) and the key results presented are therefore robust.

The distribution pattern of TP53 mutations was in line with previous studies, with the majority of mutations localized in the DNA binding domain (exons 5-8). Notably, almost 20% of the mutations were detected outside this domain, and the prognostic significance of these mutations was equivalent to those residing in the DNA-binding domain. This highlights the need to sequence all TP53 exons to explore its clinical impact.

The distribution of TP53 mutation types across its functional domains provides insight into the selection processes during tumorigenesis. Frameshift mutations, as opposed to other mutation types, were evenly distributed along the gene (Fig. 1). Thus, selective advantage of mutations in the DNA binding domain, mirroring the functional importance of this region, does not extend to frameshift mutations. This suggests that frameshift mutations disrupt protein function independently of location.

The impact of the TP53 mutation types on outcome could be related to the predicted functional effect, although reports of this association have been inconsistent(16,30-35). In this
study, the prognostic impact was not significantly different between mutation types. Missense mutations predicted to have “gain of function” (GOF)(6,36) were also not distinctly associated with outcome. The LOH events in \( TP53 \) mutated tumors were often accompanied by copy number gain (mostly duplication) of the retained allele. This occurred independently of mutation type, suggesting that allelic duplication does not enhance the effect of GOF mutant proteins. Altogether, these observations question whether different mutation types, including GOF mutations, can be used as a prognostic tool in breast cancer.

The distinct \( TP53 \) mutation spectrum and clinical relevance across molecular subtypes suggest a diverse role for TP53. As seen in previous studies(10,20), basal-like cancers were frequently mutated and enriched for frameshift and nonsense mutations, but \( TP53 \) mutation was not prognostic. In a recent study, \( TP53 \) function was found to be compromised in most of the Basal-like tumors either through \( TP53 \) mutations or alterations in genes in the TP53 pathway(10). Thus, wildtype \( TP53 \) may not necessarily indicate an intact TP53 pathway and may explain this lack of prognostic value. Our interpretation is that abrogation of the \( TP53 \) pathway is an early, initiating and required event in most Basal-like tumors. In addition, inactivation of both \( TP53 \) alleles seems to be important in these tumors given the higher proportion of cases with both mutation and LOH (Fig. S8-A). In ER negative disease, where both \( TP53 \) and other markers performed poorly in prognostication (Table S2, S3 and S4), severe lymphocytic infiltration is associated with better outcome(37-39). We observed that ER negative tumors with wildtype \( TP53 \) and severe lymphocytic infiltration are a subset with better prognosis (Fig S8-C). Whether wildtype \( TP53 \) evokes an immune response in ER-negative tumors, or whether the immune response is only effective against wildtype \( TP53 \) tumors needs to be investigated.
Recurrent mutated codons (hotspots) were observed, particularly in Basal-like/IC10 tumors. Hotspots probably arise due to a combination of a highly mutable sequence context and selective growth advantage provided by the specific mutation. One intriguing observation was the association of the nonsense mutation in codon 213 (C.R213*) with the Basal-like subtype. Codon 213 (CGA/CGG) is a polymorphic site (rs1800372) with G as the minor allele (Minor Allele Frequency=0.007)(40). The C>T mutation leading to a stop codon is possible only with the A allele (CGA>TGA), indicating a protective role of the G-allele.

Loss of one TP53 allele may be a sufficient driver in Luminal B tumors as indicated by the higher fraction of wildtype tumors with LOH (Fig. S8-A). A less important role of TP53 mutation is suggested in Luminal A tumors due to the low mutation frequency and lack of evident hotspots.

In conclusion, our integrated study provides an important insight into the significance of TP53 in breast cancer, clearly demonstrating that its role as a prognostic and predictive marker needs to be investigated in a subtype specific manner. This knowledge is crucial with respect to the potential clinical application of TP53, including further development of TP53 targeted therapies.
Authors' contributions

LSP, SFC and AL performed the sequencing analyses. LSP, HKMV and OMR performed bioinformatics analyses. SFC, TO and SMK curated and analyzed IHC and pathology data. LSP, HKMV, DQ, OMR and AL performed statistical and survival analysis. LSP, HKMV, CC, ALBD and AL wrote the paper, with input from SFC, DQ, VNK and OMR. CC, ALBD and AL conceived and directed the study. CC and SA are the principal investigators of the METABRIC study.

Acknowledgements

We thank the members of the METABRIC GROUP (Molecular Taxonomy of Breast Cancer International Consortium; list of members below) for sample collection and for providing the clinical and genomic data on 2000 breast tumors. The authors also thank the patients who contributed to the study. We wish to thank Phuong Vu, Anita Halvei, Inger Riise Bergheim, Anja Valen, Rita Halvorsen, Eldri Undlien Due, Jovana Jovanovic and Shiferaw Dagim Tadele for the efforts to read the sequences of more than 4.2 million bases manually.

List of members of the METABRIC GROUP

Co-chairs Carlos Caldas\textsuperscript{1,2}, Samuel Aparicio\textsuperscript{3,4}

Steering committee James D. Brenton\textsuperscript{1,2}, Ian Ellis\textsuperscript{5}, David Huntsman\textsuperscript{3,4}, Sarah Pinder\textsuperscript{6}, Arnie Purushotham\textsuperscript{6}, Leigh Murphy\textsuperscript{7}, Carlos Caldas\textsuperscript{1,2}, Samuel Aparicio\textsuperscript{3,4}

Tissue and clinical data source sites: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)\textsuperscript{1,2}, Helen Bardwell\textsuperscript{2}, Suet-Feung Chin\textsuperscript{1,2}, Christina Curtis\textsuperscript{1,2}, Zhihao Ding\textsuperscript{2}, Stefan Gräf\textsuperscript{2,2}, Linda Jones\textsuperscript{5}, Bin Liu\textsuperscript{1,2}, Andy G. Lynch\textsuperscript{1,2}, Irene Papaniak\textsuperscript{1,2}, Stephen J. Sammut\textsuperscript{7}, Gordon Wishart\textsuperscript{6}, British Columbia Cancer Agency Samuel Aparicio (Principal Investigator)\textsuperscript{3,4}, Steven Chia\textsuperscript{1}, Karen Gelmon\textsuperscript{2}, David Huntsman\textsuperscript{3,4}, Steven McKinney\textsuperscript{3,4}, Caroline Speers\textsuperscript{6}, Gulisa Turashvili\textsuperscript{3,4}, Peter Watson\textsuperscript{3,4}, University of Nottingham Ian Ellis (Principal Investigator)\textsuperscript{9}, Roger Blamey\textsuperscript{9}, Andrew Green\textsuperscript{9}, Douglas Macmillan\textsuperscript{9}, Emanuele di Rinaldis\textsuperscript{6}, Andy Tutt\textsuperscript{9}; King’s College London Arnie Purushotham (Principal Investigator)\textsuperscript{9}, Cheryl Gillett\textsuperscript{9}, Anita Grigoriasidis\textsuperscript{6}, Sarah Pinder\textsuperscript{6}, Emanuele di Rinaldis\textsuperscript{6}, Andy Tutt\textsuperscript{9}; Manitoba Institute of Cell Biology Leigh Murphy (Principal Investigator)\textsuperscript{7}, Michelle Parisien\textsuperscript{7}, Sandra Troup\textsuperscript{7}

Cancer genome/transcriptome characterization centres: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)\textsuperscript{1,2}, Suet-Feung Chin (Team Leader)\textsuperscript{1,2}, Derek Chan\textsuperscript{1}, Claire Fielding\textsuperscript{2}, Ana-Teresa Maia\textsuperscript{1,2}, Sarah McGuire\textsuperscript{2}, Michelle Osborne\textsuperscript{2}, Sara M. Sayalero\textsuperscript{2}, Inmaculada Spiteri\textsuperscript{2}, James Hadfield\textsuperscript{2}, British Columbia Cancer Agency Samuel Aparicio (Principal
Investigator)3,4, Gulisa Turashvili (Team Leader)3,4, Lynda Bell4, Katie Chow4, Nadia Gale4, David Huntsman3,4, Maria Kovalik3, Ying Ng4, Leah Prentice4

Data analysis subgroup: University of Cambridge/Cancer Research UK Cambridge Research Institute

Carlos Caldas (Principal Investigator)1,2, Simon Tavaré (Principal Investigator)1,2,10,11, Christina Curtis (Team Leader)1,2,*, Mark J. Dunning2, Stefan Gräf2, Andy G. Lynch1,2, Oscar M. Rueda1,2, Roslin Russell2, Shamith Samarajiw2,10, Doug Speed2,10, Florian Markowetz (Principal Investigator)1,2, Yinyin Yuan1,2, James D. Brenton (Principal Investigator)3,4, British Columbia Cancer Agency Samuel Aparicio (Principal Investigator)3,4, Sohrab P. Shah (Team Leader)3,4, Ali Bashashati3, Gavin Ha3, Gholamreza Haffari3, Steven McKimney3,4

1Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. 2Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. 3Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. 4Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. 5Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. 6King’s College London, Breakthrough Breast Cancer Research Unit, London, WC2R 2LS, UK. 7Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada. 8Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. 9Cambridge Breast Unit, Addenbrooke’s Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. 10Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK. 11Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. 1*Present address: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.
**Figure legends**

**Figure 1.** *TP53* mutation spectrum in breast cancer. **A.** Distribution of mutation type across the gene. **B.** Pie chart showing the fraction of each mutation category. **C.** Conserved and functional domains of *TP53*.

**Figure 2.** *TP53* mutation spectrum in molecular subtypes, **A.** PAM50 and **B.** IC subtypes. **C.** Pie chart showing fraction of mutation type in each subtype. **D.** Conserved and functional domains of *TP53*.

**Figure 3.** Kaplan-Meier survival curves showing BCSS of *TP53* mutation status in, **A.** the total cohort, **B.** patients categorized according to *TP53* mutation types, **C.** ER positive patients, **D.** ER negative patients. Numbers at risk are listed below each chart (e = no. of breast cancer specific deaths).

**Figure 4.** Kaplan-Meier survival curves showing BCSS of *TP53* mutation status in PAM50 and IC10 breast cancer subtypes. **A.** Luminal A. **B.** Luminal B. **C.** HER2 enriched. **D.** Normal-like. **E.** Basal-like. **F.** Integrative cluster 10 (IC10). Numbers at risk are listed below each chart (e = no. of breast cancer specific deaths).

**Figure 5.** **A.** Kaplan-Meier survival curves showing BCSS of *TP53* mutation status in combination with both *TP53* LOH and *MDM2* amplification. Numbers at risk are listed below each chart (e = no. of breast cancer specific deaths). **B.** Relation between genomic instability and TP53 pathway disrupting events (*TP53* mutation, *TP53* LOH, and *MDM2* gain).
References


35. Overgaard J, Yilmaz M, Gulberg P, Hansen LL, Alsner J. TP53 mutation is an independent prognostic


Table 1  Association between TP53 mutation status and clinicopathological characteristics (n = 1420).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>wt TP53</th>
<th>mut TP53</th>
<th>p-value***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤55</td>
<td>320</td>
<td>167</td>
<td>34.3</td>
</tr>
<tr>
<td>&gt;55</td>
<td>706</td>
<td>227</td>
<td>24.3</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ductal</td>
<td>748</td>
<td>356</td>
<td>32.2</td>
</tr>
<tr>
<td>Lobular</td>
<td>99</td>
<td>8</td>
<td>7.5</td>
</tr>
<tr>
<td>IDC+HLC</td>
<td>64</td>
<td>10</td>
<td>13.5</td>
</tr>
<tr>
<td>Tubular</td>
<td>59</td>
<td>5</td>
<td>7.8</td>
</tr>
<tr>
<td>Mucinous</td>
<td>37</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>Medullary</td>
<td>10</td>
<td>11</td>
<td>12.4</td>
</tr>
<tr>
<td>Other invasive</td>
<td>9</td>
<td>3</td>
<td>30.0</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 (≤2 cm)</td>
<td>456</td>
<td>149</td>
<td>24.6</td>
</tr>
<tr>
<td>T2 (&gt;2 ≤5 cm)</td>
<td>515</td>
<td>224</td>
<td>43.0</td>
</tr>
<tr>
<td>T3 (&gt;5 cm), T4</td>
<td>51</td>
<td>18</td>
<td>26.0</td>
</tr>
<tr>
<td>Missing</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Node status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg (pN0)</td>
<td>539</td>
<td>177</td>
<td>24.7</td>
</tr>
<tr>
<td>Pos (pN1-pN3)</td>
<td>483</td>
<td>216</td>
<td>30.9</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>119</td>
<td>6</td>
<td>4.8</td>
</tr>
<tr>
<td>2</td>
<td>479</td>
<td>61</td>
<td>13.3</td>
</tr>
<tr>
<td>3</td>
<td>383</td>
<td>320</td>
<td>45.5</td>
</tr>
<tr>
<td>Missing</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrogen receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pos</td>
<td>906</td>
<td>195</td>
<td>17.7</td>
</tr>
<tr>
<td>Neg</td>
<td>120</td>
<td>199</td>
<td>62.4</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progesteron receptor*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pos</td>
<td>627</td>
<td>116</td>
<td>15.6</td>
</tr>
<tr>
<td>Neg</td>
<td>399</td>
<td>278</td>
<td>41.1</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HER2*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>184</td>
<td>134</td>
<td>42.1</td>
</tr>
<tr>
<td>Loss/balanced</td>
<td>840</td>
<td>257</td>
<td>23.4</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP53 LOH**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>403</td>
<td>312</td>
<td>43.6</td>
</tr>
<tr>
<td>No</td>
<td>611</td>
<td>74</td>
<td>10.8</td>
</tr>
<tr>
<td>Missing</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDM2 amplification**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>311</td>
<td>194</td>
<td>38.4</td>
</tr>
<tr>
<td>No</td>
<td>693</td>
<td>169</td>
<td>19.6</td>
</tr>
<tr>
<td>Missing</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAM50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal-like</td>
<td>82</td>
<td>152</td>
<td>65.0</td>
</tr>
<tr>
<td>ERBB2*</td>
<td>75</td>
<td>86</td>
<td>53.4</td>
</tr>
<tr>
<td>Luminal B</td>
<td>285</td>
<td>94</td>
<td>24.8</td>
</tr>
<tr>
<td>Luminal A</td>
<td>460</td>
<td>47</td>
<td>9.3</td>
</tr>
<tr>
<td>Normal-like</td>
<td>121</td>
<td>15</td>
<td>11.0</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrative Cluster (IC10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC1</td>
<td>83</td>
<td>24</td>
<td>22.4</td>
</tr>
<tr>
<td>IC2</td>
<td>43</td>
<td>11</td>
<td>20.4</td>
</tr>
<tr>
<td>IC3</td>
<td>192</td>
<td>13</td>
<td>6.3</td>
</tr>
<tr>
<td>IC4</td>
<td>207</td>
<td>42</td>
<td>16.9</td>
</tr>
<tr>
<td>IC5</td>
<td>64</td>
<td>74</td>
<td>53.6</td>
</tr>
<tr>
<td>IC6</td>
<td>41</td>
<td>24</td>
<td>36.9</td>
</tr>
<tr>
<td>IC7</td>
<td>114</td>
<td>17</td>
<td>13.0</td>
</tr>
<tr>
<td>IC8</td>
<td>191</td>
<td>14</td>
<td>7.8</td>
</tr>
<tr>
<td>IC9</td>
<td>52</td>
<td>48</td>
<td>48.0</td>
</tr>
<tr>
<td>IC10</td>
<td>39</td>
<td>127</td>
<td>76.5</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjuvant therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>9</td>
<td>22</td>
<td>71.0</td>
</tr>
<tr>
<td>CT/RT</td>
<td>45</td>
<td>86</td>
<td>65.6</td>
</tr>
<tr>
<td>CT/HT/RT</td>
<td>99</td>
<td>43</td>
<td>30.3</td>
</tr>
<tr>
<td>CT/HT</td>
<td>19</td>
<td>7</td>
<td>26.9</td>
</tr>
<tr>
<td>HT</td>
<td>245</td>
<td>56</td>
<td>18.6</td>
</tr>
<tr>
<td>HT/RT</td>
<td>375</td>
<td>100</td>
<td>21.1</td>
</tr>
<tr>
<td>RT</td>
<td>110</td>
<td>43</td>
<td>38.1</td>
</tr>
<tr>
<td>None</td>
<td>124</td>
<td>37</td>
<td>23.0</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Progesteron status defined by gene expression data. HER2 status defined by SNP6 data.
** TP53 LOH and MDM2 gain defined by SNP6 data (not corrected for ploidy).
*** Pearson Chi-Square or Fisher's exact test (2-tailed) when 2x2.
Figure 3

A. All patients (n=1404)
   TP53 wildtype (e=233)
   TP53 mutant (e=158)
   p < 0.001

B. TP53 mutation types (n = 393)
   Non-missense (e=88)
   Missense non-DBM (e=26)
   Missense DBM (e=64)
   p = 0.454

C. ER positive patients (n = 1090)
   TP53 wildtype (e=188)
   TP53 mutant (e=80)
   p < 0.001

D. ER negative patients (n = 314)
   TP53 wildtype (e=45)
   TP53 mutant (e=78)
   p = 0.622
**Figure 4**

A. Luminal A (n = 499)

- TP53 wildtype (e=71)
- TP53 mutant (e=8)

B. Luminal B (n = 375)

- TP53 wildtype (e=85)
- TP53 mutant (e=41)

c. HER2 enriched (n = 158)

- TP53 wildtype (e=27)
- TP53 mutant (e=44)

D. Normal-like (n = 136)

- TP53 wildtype (e=24)

E. Basal-like (n = 233)

- TP53 wildtype (e=25)
- TP53 mutant (e=56)

F. IC 10 (n = 166)

- TP53 wildtype (e=12)
- TP53 mutant (e=39)

*Author Manuscript Published OnlineFirst on May 6, 2014; DOI: 10.1158/1078-0432.CCR-13-2943*

*Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.*
Figure 5

A. TP53 mut_LOH_MDM2 ampl (n = 1352)

- 0 events (e=73)
- 1 event (e=118)
- 2 events (e=123)
- 3 events (e=65)

p < 0.001

B. Events abrogating the TP53 pathway

Genetic Instability Index

p < 0.001
TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance

Laxmi Silwal-Pandit, Hans Kristian Moen Vollan, Suet-Feung Chin, et al.

Clin Cancer Res  Published OnlineFirst May 6, 2014.

Updated version  Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-2943

Supplementary Material  Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/05/07/1078-0432.CCR-13-2943.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.