DNA methylation profiling to assess pathogenicity of BRCA1 unclassified variants in breast cancer

Kirsty J. Flower¹, Natalie S. Shenker¹, Mona El-Bahrawy², David E Goldgar³, Michael T. Parsons⁴, KConFab Investigators⁵, AFFECT study group⁶, Amanda B. Spurdl⁻, Joanna R. Morris⁶, Robert Brown¹,⁷ and James M. Flanagan¹

¹Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, UK
²Department of Histopathology, Hammersmith Hospital, Imperial College London, UK
³Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
⁴QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
⁵Peter MacCallum Cancer Center, Melbourne, VIC, Australia
⁶Genome Stability Unit, School of Cancer Sciences, University of Birmingham, UK
⁷Section of Molecular Pathology, Institute for Cancer Research, Sutton, UK

*Address for correspondence:

Dr. James M. Flanagan
Epigenetics Unit, Division of Cancer, Department of Surgery and Cancer,
Faculty of Medicine, Imperial College London
4th Floor IRDB, Hammersmith Campus, Du Cane Road, London W12 0NN
Tel: 020 75942127; fax: 020 75942129; email: j.flanagan@imperial.ac.uk
Abstract

Germline pathogenic mutations in BRCA1 increase risk of developing breast cancer. Screening for mutations in BRCA1 frequently identifies sequence variants of unknown pathogenicity and recent work has aimed to develop methods for determining pathogenicity. We previously observed that tumour DNA methylation can differentiate BRCA1 mutated from BRCA1 wild type tumours. We hypothesised that we could predict pathogenicity of variants based on DNA methylation profiles of tumours that had arisen in carriers of unclassified variants. We selected 150 FFPE breast tumour DNA samples (47 BRCA1 pathogenic mutation carriers, 65 BRCAx (BRCA1 wild-type), 38 BRCA1 test variants) and analysed a subset (n=54) using the Illumina 450K methylation platform, using the remaining samples for bisulphite pyrosequencing validation. Three validated markers (BACH2, C8orf31 and LOC654342) were combined with sequence bioinformatics in a model to predict pathogenicity of 27 variants (independent test set). Predictions were compared with standard multifactorial likelihood analysis. Prediction was consistent for c.5194-12G>A (IVS 19-12 G>A) (P>0.99); 13 variants were considered not pathogenic or likely not pathogenic using both approaches. We conclude that tumour DNA methylation data alone has potential to be used in prediction of BRCA1 variant pathogenicity but is not independent of ER status and grade which are used in current multifactorial models to predict pathogenicity.

Keywords

Methylation, Breast cancer, BRCA1, epigenetic, variants
Introduction

Breast cancer remains the most diagnosed cancer in women, with an overall incidence in the UK of 1 in 8 (http://www.cancerresearchuk.org/cancer-info/cancerstats/) but fortunately advances in treatment and screening have resulted in reduced mortality rate ¹. We now understand that breast cancer is a heterogeneous disease, with different individuals benefiting from different treatments depending on hormone receptor expression and genetic mutations. Distinct novel molecular subgroups have been identified using a wide array of technologies, suggesting the need for novel therapies targeting specific molecular alterations ², ³. A minority of breast cancers (5-10%) are considered hereditary, and approximately a quarter of these are due to germline mutations in known cancer susceptibilities genes, including \textit{BRCA1} (OMIM# 113705) and \textit{BRCA2} (OMIM# 600185) ⁴. Families presenting with multiple breast cancer cases, and families with early onset ovarian cancer, are currently tested by sequencing these two genes, and this informs clinicians in the management of individual treatment, and decisions concerning screening, chemoprevention or prophylactic surgery for unaffected mutation carriers. Both triple negative tumours and \textit{BRCA1} germline mutated tumours are significantly enriched in the basal subtype, which is associated with poor prognosis. Current estimates suggest that up to 15% of triple negative tumours are \textit{BRCA1} germline mutation carriers ⁵, ⁶.

\textit{BRCA1} variants that are considered pathogenic mutations include those leading to protein truncation or mRNA degradation, and risk-associated missense alterations affecting protein function accompanied by clinical evidence of pathogenicity. However, there are also a considerable number of \textit{BRCA1} variants of uncertain clinical significance (>500 different variants ⁷), namely missense variants with undetermined effect on function or risk, small insertions or deletions, or alterations in non-coding sequence. Attempts to classify these unclassified variants have utilised a range of analyses. Segregation data can be used to identify if a variant tracks with disease within a family, however this is difficult to apply to some variants since they are individually rare ⁸. \textit{In silico} approaches have been used to examine sequence evolutionary conservation predicting the effect of specific amino acid substitutions ⁹ and potential splicing alterations ¹⁰. A multifactorial model has been developed to integrate these different data together, starting with an empirical prior probability based on bioinformatic predictions, and incorporating likelihood ratios derived from independent data sources to generate a posterior probability of pathogenicity ⁸. The advantages of using this approach is that it incorporates all available information from multiple data types in a single analysis with a numeric output, and allows for the addition of extra data ¹¹. The posterior probability generated can then be used to classify the variant into 1 of 5
classes: Class 1 (probability <0.001) - not pathogenic, or of no clinical significance; Class 2 (probability 0.001-0.049) – likely not pathogenic or of little clinical significance; Class 3 (probability 0.05-0.949) – uncertain; Class 4 (probability 0.95-0.99) – likely pathogenic; Class 5 (probability >0.99) – definitely pathogenic 12.

We have previously shown that breast tumours from patients with pathogenic BRCA1 germline mutations have distinct DNA methylation profiles compared to familial breast cancer cases with no BRCA1 or BRCA2 mutations (BRCAx) 13. In this previous study, 81.3% tumours with germline mutated BRCA1 were correctly predicted using a support vector learning machine approach based on DNA methylation data. The methylation profiles of BRCAx tumours are very similar to that of sporadic breast cancers 14. The present study aimed to assess the DNA methylation of both published candidate genes and novel regions that differ between tumours from BRCA1 germline mutation carriers (referred to as BRCA1 tumours) and tumours from BRCAx families (BRCA1 germline wildtype), with a view to providing an additional factor useful for classification of BRCA1 unclassified variants.
Results

DNA methylation of prior candidate genes is dependent on ER status

Candidate regions were identified from previous data 13, 14 where DNA methylation levels were significantly different between tumours of BRCA1 mutation carriers and those of BRCAx individuals (Supp. Table S1). Twelve candidate regions were analysed in bisulphite converted DNA from 150 tumours (BRCA1, BRCAx and BRCA1 test variant) by pyrosequencing. We validated the expected difference between BRCA1 and BRCAx tumours for 6/12 candidate regions, observing lower median methylation in BRCA1 tumours for five candidate regions (CD9 (OMIM# 143030), SGK1 (OMIM# 602958), ERCC3 (OMIM# 133510) and 2 regions in FGF2 (OMIM# 134920), wilcoxon rank sum test, p<0.05) and significantly higher methylation in BRCA1 tumours for one gene (CD40 (OMIM# 109535), p=0.02) (Supp. Table S1). Logistic regression analysis showed that methylation status was significantly associated with BRCA1 mutation status (Supp Table S1). Since ER status is a known predictor of BRCA1 mutation status (LR ranging from 0.08-0.90 for ER-negative status, dependent on grade, 15), we assessed whether differences observed in methylation between BRCA1 and BRCAx tumours in the remaining five candidate regions of interest were independent of these known histopathological predictors. Using a generalised linear model we showed that methylation differences observed in these five regions were associated with ER status (p value <0.05) (Supp. Table S1). ER status and mutation status are highly associated in this dataset (glm, p-value= 4.31e-7), and none of the five candidates were significantly independent from ER status, indicating that they should not be used in conjunction with existing pathology LRs 15 in multifactorial modelling for prediction of variant pathogenicity. In light of these findings, further analyses considering grade, another predictor of mutation status, were not pursued.

Genome-wide analysis identifies novel differentially methylated loci that define BRCA1 mutation status

The overall study design is shown in Figure 1A and 1B. The Illumina 450K beadchip array has successfully been used to assess DNA methylation in FFPE derived DNA samples using appropriate quality control and restoration methods 16. We selected 60 samples (BRCA1 pathogenic variants n=20, BRCAx n=20, BRCA1 test variant n=20) for array analysis, of which 54 samples passed quality control (90%: BRCA1 pathogenic n=18, BRCAx n=19, BRCA1 test variant n=17) and 482351/485577 probes (99.3%) passed QC. The Wilcoxon rank sum test was used in three analyses to identify (i) probes with significant differences between tumours from patients with germline BRCA1 mutations and tumours from BRCAx cases (fdr q<0.05, n=250 probes) (ii) probes with significant differences between ER positive tumours and ER negative tumours (fdr q<0.05,
n=55148 probes) and (iii) probes with significant differences between low grade (1 and 2) and high grade (grade 3) tumours (fdr q<0.05, n=29 probes). Using a Venn diagram we show that there were 23 probes were apparently unique to the mutation class analysis (Figure 2A). Probes with an absolute methylation difference between groups of less than 5% were excluded, leaving 18 probes with a range of absolute difference in methylation between 5% and 30%. We used consensus clustering of the array data using the 18 selected probes that differentiated BRCA1 pathogenic tumours from the BRCAx tumours to determine which cluster each of the BRCA1 test variant tumours clustered with. This method confirmed only two tumour clusters based on these data (Figure 2B) and found nine BRCA1 test set variant tumours clustered with pathogenic BRCA1 mutated tumours, and eight BRCA1 test set variant tumours clustered with BRCAx tumours (Figure 2C).

Pyrosequencing assays were designed for these 18 loci, and optimized for 16. Methylation analysis of these regions was conducted for 150 FFPE samples (54 matched to the array, the 90 independent samples, and the 6 that did not pass array QC criteria). Intraclass correlation coefficients (ICC) of assays were calculated to compare the beta values obtained from the array to pyrosequencing values for matched samples, and the Wilcoxon rank sum test used to validate the difference between the BRCA1 mutated and BRCAx group in the independent samples (Table 1). A significant difference (p<0.05) between BRCA1 samples and BRCAx samples was validated for 5 loci (450K probe IDs cg24667115 (BACH2 OMIM# 605394), cg03029255 (C8orf31), cg02502358 (LYRM9), cg21645762 (LOC654342) and cg12472473 (chr13q34)) in the independent group of samples.

However, the cg12472473 (13q34) locus harboured a CG>TG SNP with a minor allele frequency (MAF) of 0.2 within the CpG dinucleotide of interest, and pyrosequencing genotyping showed that the T genotype was overrepresented in the BRCAx group (chi squared test, p<0.05). This region was therefore excluded from further analysis. The technical and biological validation of the four loci is shown in figure 3. Further, despite selecting array probes that apparently predicted of mutation status independent of ER and grade status (see Figure 2), logistic regression analysis of the pyrosequencing data showed that probes cg24667115 (BACH2) and cg02502358 (LYRM9) were not independently associated with mutation status when ER status was also included in the model (Table 2). Additionally, there was evidence that methylation status of cg02502358 (LYRM9), cg24667115 (BACH2) and cg21645762 (LOC654342) were correlated with grade, although this analysis was
based on approximately half of the individuals, for which we had information on all variables, which reduces the statistical power of the analysis.

As for candidate region analysis described above, these findings suggest that the majority of methylation probes that are associated with mutation status are, in fact, not independent of ER or grade. Thus, with the exception of cg03029255 (C8orf31), the markers identified in this study should not be used in conjunction with existing ER and grade LRs in multifactorial likelihood prediction of variant pathogenicity.

**Predictive capacity of different methylation markers of mutation status**

Pearson’s correlation coefficient was calculated for the pyrosequencing data of each locus against the others, and the $R^2$ value for all comparisons was less than 0.2, indicating a low extent of correlation ($BACH2$ vs $C8orf31 = 0.0436$, $BACH2$ vs $LOC654342 = 0.1719$, $BACH2$ vs $LYRM9 = 0.0207$, $C8orf31$ vs $LOC654342 = 0.1935$, $C8orf31$ vs $LYRM9 = 0.0953$, $LOC654342$ vs $LYRM9 = 0.0356$).

The beta values at these four loci were converted to z-scores and combined in a logistic regression model. Due to missing data for some of these markers, there were not enough data points to fit a model to more than three markers. $LYRM9$ was therefore removed from the model. Using the Leave One Out Cross Validation (LOOCV) method based on methylation data of the three remaining markers alone, 32 out of 37 samples are correctly predicted (86%). The pyrosequencing methylation data was also converted to z-scores, and the combined logistic regression model correctly predicted 78 out of 97 (80% - PPV=76%, NPV=79%) of samples with known mutation status using only methylation data.

The logistic regression model including methylation of the three validated loci was used to predict all tumour samples from test set variant carriers based on methylation alone, and results were compared to current classification based on other evidence including ER and grade among other variables (Table 3). To represent these predictions for all samples based on their methylation, included those of known pathogenicity, likelihood ratios (mLRs) were calculated using the probability values generated by the prediction model (probability of pathogenicity divided by 1 minus the probability of pathogenicity, shown in Supp. Tables S2 and S3), and the log of these values was plotted (Figure 4). This indicates which samples of the same variant were similar in pathogenicity prediction (for example, $BRCA1$ IVS 19-12 G>A) and those which are more discordant.
(Arg1347Gly), and also illustrates the spread of methylation prediction for the known variants (BRCA1 and BRCAx extremes of plot).

Of the 27 unique variants in the test set, one is known to be an intermediate risk allele from extensive segregation studies (Arg1699Gln), and 15 have sufficient information available to place them in class 1,2, or 4,5 based on multifactorial likelihood analysis that uses currently accepted predictors of mutation status. Of the latter, one is considered class 5 (pathogenic), one is class 4 (likely pathogenic), seven are class 1 (not pathogenic), 6 are class 2 (likely not pathogenic), and the remainder remain class 3 (uncertain) after classification. Comparing these segregation and multifactorial results to those from methylation analysis alone, the intermediate risk variant Arg1699Gln had an LR of 25.218, the two class 5 variant IV19-12 G>A samples had LRs of 19.6 and 27.24 in favour of pathogenicity, and the class 4 variant BRCA1 IVS14+2 ins8 had an LR of 1.43. The class 1 variants variously had LRs ranging from 0.008 to 19.75 (with more than 60% of these <1), while the class 2 variants had LRs ranging from 0.02 to 1.05.

There were 27 unique BRCA1 variants within these 37 individuals, five of which had at least two independent tumour samples from different individuals analysed, and the logistic regression model results in probabilities of pathogenicity for each sample. These probabilities were converted to LRs and combined by multiplication. Prior probabilities for these variants based on bioinformatic analysis (see methods) were combined with the likelihood ratio calculated for each variant, and these posterior probabilities are summarised in table 3. Using the established posterior probability cut-offs for the IARC 5 tier classification system, BRCA1 IVS 19-12 G>A would be classified as pathogenic (class 5). Arg496Cys and Arg841Trp would be classified as neutral (class 1). This would suggest that, at least based on current information, the likelihood ratios derived from the methylation results based on three probes are concordant for the majority of variants with the largest discordance observed being with class 3 (uncertain) variants. When these likelihood ratios are combined with the bioinformatic prior probabilities, the predictions for class 1 and 2 variants become far more consistent, but the prediction for BRCA1 IVS14+2 ins8 becomes more uncertain.
Discussion

Multiple studies have identified distinct epigenetic profiles that correlate with different breast cancer subtypes and/or ER status. The most recent example is data from The Cancer Genome Atlas (TCGA) project which used shared probes between 27K and 450K analysis on 802 tumours to distinguish five groups. Two groups were highlighted: group 3 showed a hypermethylated phenotype and was enriched for luminal subtype ER positive tumours, whilst group 5 showed the lowest levels of methylation in the probes selected and this group overlapped with basal-like subtype ER negative tumours. The association between methylation and subtype/ER status is recapitulated by several other studies, which show a lower level of methylation in ER negative/basal-like tumours compared to ER positive/luminal tumours. We previously showed a difference in methylation between familial breast tumours with different BRCA1 germline mutations compared to BRCAx, and found that the methylation clusters formed were independent of the intrinsic subtype, and therefore had the potential to be independent of ER status. We hypothesised that this difference in DNA methylation profile could contribute evidence to classify BRCA1 sequence variants of uncertain clinical significance.

We undertook a two-phase study to assess the value of DNA methylation for predicting pathogenicity of BRCA1 variants, the strength of which was the large number of BRCA1 mutant carrier and BRCAx (BRCA1 wildtype) tumours used as a reference to generate the predictive algorithm. There is no evidence to suggest that BRCAx tumours have significantly different methylation profiles to sporadic breast cancers. The first phase analysed 12 previously reported candidate regions and validated five regions to be strongly associated with mutation status, but also showed that these associations were not independent of ER status or grade. In the second phase using the Illumina 450K beadchip array to conduct a genome-wide analysis, we showed that DNA methylation profiles are largely driven by ER status, with 55148 probes significantly associated with ER status. The mechanism by which this occurs is still not understood, and it is unknown whether these changes in methylation occur at ER target genes or as a consequence of ER transcriptional regulation. Very few probes were significantly associated with grade.

Using the Illumina 450K beadchip array, we also identified 250 loci associated with mutation status, 23 of which initially appeared to be independent of tumour ER status and grade. Of these, 18 had an absolute methylation difference between groups of greater than 5%, and using these 18 novel loci, we were able to
cluster the samples into two groups of *BRCA1* carrier and *BRCAX* tumours. The test set samples from variant carriers clustered with the *BRCA1* pathogenic or *BRCAX* sub-groups. Of these 18 loci, four sites were validated by pyrosequencing in an independent group of samples, but logistic regression suggested that for all but one probe, the association was not truly independent from ER status and grade in the independent samples. Using the data to predict the mutation status from these methylation values alone, a logistic regression model was created and used to predict the pathogenicity of the 27 different test variants, and these predictions based on methylation data were compared to current evidence regarding pathogenicity using other information sources (summarised in Table 3, detailed in Supp table 4). Combining the predictions from the methylation models with known prior probabilities for each variant, *BRCA1* IVS 19-12 G>A would be classified as pathogenic (Class 5), and Arg496Cys and Arg841Trp would be classified as neutral (Class 1). These classifications matched those based on multifactorial likelihood analysis or segregation analysis, so the utilisation of this methylation data supports the predictions of these variants.

Further comparison of the current classifications and the methylation derived classifications revealed very few major inconsistencies; a further 8 variants (Ala1368Val, Gly275Ser, Arg1347Gly, His971Arg, *BRCA1* IVS 12+9 C>T, Pro346Ser, Pro727Leu, Ser475Cys) were classified by both methods as class 1 (not pathogenic) or class 2 (likely not pathogenic). A further three variants (Ser1655Pro, *BRCA1* IVS 23+5 G>C, *BRCA1* IVS 9+16 C>G) were classified by both methods in the broad uncertain class 3, indicating for these few variants there is not enough evidence to categorise them confidently by either method. The methylation derived classification correctly predicts the class 4 variant within this test set, *BRCA1* IVS 14+2 ins 8, as class 4. Due to missing pathology, lack of previous investigation or a combined LR between 0.5 and 2.0 (which therefore does not pass thresholds recommended by the ENIGMA BRCA classification guidelines, see http://www.enigmaconsortium.org/), six variants do not have a current classification based on multifactorial likelihood analysis or segregation analysis but are all classified by the methylation model as class 1 or class 2 (Ile68Arg, Leu152Phe, *BRCA1* IVS 9+16C>G, Asn1236Lys, Arg1762Gly, Ser475Cys). It is therefore difficult to assess the validity of the methylation predictions for these variants. Finally, two variants with current uncertain classification (Asp214Gly and His662-Arg664del) share methylation characteristics with the less pathogenic variants of class 1 and 2.
The relationship between loss of BRCA1 function and a direct or indirect mechanism that influences DNA methylation in normal or cancerous breast tissue remains unclear and an area for future research. We observed two samples that exhibited BRCA1 promoter methylation (>10% by pyrosequencing); one BRCAx sample and one test variant sample (BRCA1 IVS 9+16C>G). The BRCAx sample was predicted to be more similar to a pathogenic BRCA1 mutation (probability from meth model = 0.8778, LR=7.18) (Supp Table 3), while the test variant also had a probability of 0.89 and LR of 8.1575 (Supp Table 2), indicating the potential similarity with pathogenic variants, however, the posterior probability for this sample was 0.14 and remained class 3 (uncertain). Therefore, it remains unclear whether BRCA1 promoter methylation, which may be more heterogeneous in the different tumour cells, influences the DNA methylation profile in the same manner as germline BRCA1 pathogenic mutations.

The BRCA1 variant Arg1699Gln has been shown to be defective in the formation of foci in response to DNA damage, and also have some effect on transcriptional activity, but that this variant was not categorised as high risk, due to the mixture of intermediate and defective phenotypes in the functional assays\textsuperscript{25}. Another study, using clinical parameters to assess pathogenicity, classified this variant as deleterious\textsuperscript{26}, and a structural approach also indicated a pathogenic phenotype\textsuperscript{27}. A recent functional complementation assay\textsuperscript{28} provided additional evidence of pathogenicity, based on proliferation and cisplatin response assays, as well as sensitivity to PARP inhibitors. The most extensive genetic study of 68 families showed that this variant was associated with an intermediate risk\textsuperscript{17}. Our logistic regression model based on DNA methylation indicated that this variant was probably pathogenic, returning a probability of pathogenicity of 0.97998. The methylation data supports increased risk known to be associated with this variant.

One of the limitations of our study is the small number of ER positive tumours with BRCA1 pathogenic mutations, and ER negative tumours without any known mutations (BRCAx). However, this is indicative of an existing bias, which makes it difficult to discover mutation specific changes, when the ER status is so closely related to the methylation levels, and this bias is inherent in most available study data. An interesting question to consider is whether the ER positive BRCA1 tumours are in fact sporadic cases; recent studies, including whole genome massively parallel sequencing analysis of both ER positive and negative BRCA1 mutated tumours, provide evidence that this is not the case\textsuperscript{29,30}. A further limitation of this study is the assumption, based on previous data\textsuperscript{13}, that all BRCA1 mutant tumours have similar profiles for different missense or truncating
mutations and the underlying hypothesis that aberrant BRCA1 function is the driver of the aberrant methylation profile. Much larger numbers of independent tumours with the same BRCA1 variants would be required to address this limitation. A limitation of using pyrosequencing is that the fragmented DNA from FFPE samples resulted in higher numbers of samples failing QC in each of the assays. Array based methods have restoration procedures which are unsuitable for pyrosequencing, and thus we found between 30%-75% of samples variably failed QC for each assay (Table 1). Due to the relative rarity of these variants, FFPE blocks are the most available source of material, however they are subject to degradation in the quality of DNA, as well as lower yields depending on size of tumour and amount of remaining archival tissue. Furthermore, we did not have access to adjacent normal tissue of the same patients which may have provided further information on whether the genetic variants influence the methylation of normal tissue in addition to the tumour tissue. Lastly, common to many tumour methylation profiling studies, we have not accounted for the numerous environmental exposures and factors that might influence or confound the methylation profile of the tumour, such as age, BMI, alcohol, smoking and the tumour microenvironment. Future studies should record such data on subjects to allow adjusting for these factors in defining differential methylation.

In conclusion, we have developed a methylation based prediction tool that adds useful information that may be included in the multifactorial model to classify BRCA1 variants of uncertain clinical significance, but note that most methylation markers identified are not truly independent of ER and grade status. Thus methylation data should be considered as complementary to other pathology data for multifactorial likelihood prediction modelling and may be useful for classifying variants with tumours where these clinical variables may be absent. Measurement of methylation allows a quantitative approach, and with the appropriate controls methylation characterisation provides promise to reduce inter-laboratory variability in pathological marker calls (such as immunohistochemistry methods 31-33), and improve application of this alternative tumour pathology characteristic for mutation prediction studies in the future. The methylation data measured using pyrosequencing can be incorporated into prediction models easily, and may capture risk information associated with multiple other pathological features. High throughput analysis of multiple samples is also feasible, and we have shown here that FFPE-derived DNA is amenable to this analysis, allowing the use of archival tissue, essential for the investigation of rare variants. In summary, this work suggests that the methylation markers will have value for future variant classification for BRCA1 and potentially for other genes with known tumour methylation phenotypes, such as MLH1 (OMIM# 120436) in colon cancer (Cancer Genome Atlas Network, 2012b).
**Materials and Methods**

**Samples**

A total of 150 breast tumour DNA samples were available for analysis, comprising of three groups: *BRCA1* germline mutated tumours (henceforth referred to as *BRCA1* n=47), *BRCA1* germline wildtype tumours from women from high risk families (BRCAx, n=65) and the designated test variant tumours (n=38). 37 of these samples (*BRCA1* n=14, BRCAx n=23) were previously extracted and described. 90 samples of known *BRCA1/2* germline status (BRCAx n=43, *BRCA1* n=32, *BRCA1* test variants n=15) were collected by the Kathleen Cunningham Foundation for Research into Breast Cancer (kConFab) consortium and ethical approval for recruitment was obtained from the institutional review boards or ethic committees of all sites, and written informed consent was given by each participant. The tumours designated as BRCAx came from women from high-risk families ascertained by kConFab, and in each case the tumour donor had undergone germline *BRCA1/2* mutation testing by full sequencing of the coding region and splice junctions and multiplex ligation dependent probe amplification (http://www.kconfab.org/). Additional tumours from *BRCA1* test variant carriers were collected by the AFFECT study (n=23), for which ethics approval for recruitment was obtained from Brighton East Ethics Committee (REC: 06/Q1907/135) and each participant gave written informed consent. Although test variants were considered unclassified at study initiation, additional information has since allowed class to be assigned to at least some of these (See below). All variants are described using the cDNA nucleotide numbering which uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. A representative section of each FFPE tumour samples was stained by H&E and evaluated by a pathologist to verify tumour content (>70% tumour) and histology, and between 3 and 9 unstained slides from each tumour were needle-macro dissected, before standard phenol chloroform DNA extraction and ethanol precipitation. The locations of the exonic sequence variants (pathogenic and test variants) assessed within *BRCA1* gene are represented in Figure 1A (6 additional intronic variations investigated are not shown). Several variants had tumour samples from more than one carrier; seven of the designated test set (table 3) and nine of the pathogenic variant set (Supp. Table S3). This study was approved by South West London REC4 (REC: 11/LO/0145).
Laboratory Analysis

Analysis of candidate regions:
Candidate regions were identified previously and in addition using likelihood ratio analysis on publicly available Illumina Goldengate array data for BRCA1 and BRCAx breast cancer tumours. All pyrosequencing assays were designed using the PyroMark Assay Design software. A common tag was placed on either the forward or reverse primer (depending on the strand to be sequenced) and a common universal biotinylated primer was used for all reactions in a semi-nested two round PCR assay. PCR primers, cycling conditions, sequencing primers and sequence to analyse are detailed in Supp. Table S5. Assays were optimised with fully methylated gDNA (100%) (Zymo Research) compared to unmethylated DNA (0%, whole genome amplified DNA (GenomiPhi V2, GE Healthcare). All PCR products were confirmed to be single bands by agarose gel electrophoresis. Assay quality was further assessed by comparing matched FFPE and fresh frozen samples with correlation coefficients ranging from R-squared=0.65 to 0.83 and agreement between array and pyrosequencing measured using intraclass correlation (table1). Methylation values were calculated as an average of all CpG sites within each assay, as determined by the Pyro-QCpG software (Qiagen).

Illumina 450K BeadChip genome-wide analysis:
Array sample selection was based upon DNA concentration quantified by Picogreen fluorescent nucleic acid stain (Invitrogen), qPCR quality control performance (Illumina FFPE QC kit) and multiplex GAPDH PCR. Bisulphite conversion of 500 ng of each sample was performed using the EZ-96 DNA Methylation-Gold™ Kit according to the manufacturer’s protocol (Zymo Research, Orange, CA). Samples underwent restoration using the Illumina Infinium HD Restoration protocol, and 4 μl of bisulphite-converted restored DNA was used for hybridization on the Infinium HumanMethylation 450 BeadChip, following the Illumina Infinium HD Methylation protocol. Hybridisation, scanning and raw data processing was performed by UCL Genomics (www.genomics.ucl.ac.uk). The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software, which normalizes within-sample data using different internal controls that are present on the HumanMethylation 450 BeadChip and internal background probes. The methylation score for each CpG was represented as a β-value according to the fluorescent intensity ratio representing any value between 0 (unmethylated) and 1 (completely methylated). Raw microarray data and processed normalized data will be available from Gene Expression Omnibus (GEO) (GSE72277). Post-array sample QC was implemented.
to check technical aspects of the array (staining, hybridisation, target removal, extension, bisulphite conversion, specificity, non-polymorphism and negative controls) using a threshold of $>3SD$ outside the normal distribution of each probe set across all samples. In total, 54 samples (90%: BRCA1 n=18, BRCAx n=19, BRCA1 test variant n=17) passed these assessments. Detection p-values were also used to evaluate probe performance, and those probes which failed in more than 20% of samples were discarded, leaving 482351/485577 probes (99.3%) available for analysis. Peak-based correction was used to normalise the data between the two probe types, and COMBAT was used to correct for batch effect between chips.

**Statistical Analysis**

Wilcoxon rank sum test was used to determine statistical significance between methylation data (pyrosequencing percentage or 450K array beta value) of the BRCA1 and BRCAx group, with FDR adjusted p<0.05 considered significant. A generalised linear model (glm), implemented in R, was used to interrogate the independence of mutation status, ER status and grade, and to generate a prediction model, which was utilised by the predict command in R for predicting variant pathogenicity. The clustcomp command from the clusterCons R package was used to perform consensus clustering on beta values of the 18 probes of all samples on the Illumina 450K beadchip array. Pearson’s correlation coefficient between probes was calculated using the cor function in R, and reported as an $R^2$ value.

A generalised linear model was constructed and used to predict pathogenicity using the “predict” function in R version 2.15.1. The resulting probabilities were converted to likelihood ratios by calculating probability/1-probability. The posterior probability was calculated as followed: the posterior odds was calculated using prior probabilities obtained from sequence bioinformatics (as described below) multiplied by the likelihood ratio of each variant and 1/1-prior probability. This was then converted to a poster probability by dividing this value by itself plus 1.

**Assessment of test set variant classifications based on current evidence**

After completion of analysis and during manuscript preparation, an extensive literature review and database search was undertaken to identify the most up-to-date information pertinent to variant classification for all the test set variants. The most up-to-date posterior probability based on multifactorial likelihood analysis was accessed from a public website displaying information collated from the literature (http://brca.iarc.fr/LOVD/home.php). For variants that did not reach class 1 (not pathogenic or of little clinical
significance) or class 5 (pathogenic), the posterior probability was recalculated using additional information relevant to multifactorial analysis. Methods used were as described previously, with two exceptions. Firstly, prior probability of pathogenicity was updated to incorporate possible effects of sequence variation of splicing, based on in silico splicing prediction algorithms adapted for this purpose (Tavtigian, personal communication). Specifically, all exonic sequence variants, plus intronic variants detected in the vicinity of the splice junction sequences, with allele frequencies <0.5%, were scored for their potential impact on splicing using MaxEntScan (MES), which computes the maximum entropy score of a given sequence using splice site models trained on human data. MES was calibrated by calculating the average and standard deviation of MES scores for the wild-type splice junctions in BRCA1, BRCA2, and ATM, allowing raw MES scores to be converted into z-scores. Based on BRCA1 and BRCA2 mutation screening data used previously to calibrate Align-GVGD rare variants that fall within the acceptor or donor region and reduce the MES score for the splice signal in which they fall showed ~97% probability to damage splice junction function when they result in a calibrated MES score of z<-2 (for donors) or z<-1.5 (for acceptors), or ~34% probability when they result in a calibrated MES score of -2<z≤0 (for donors) or -1.5<z≤+0.5 (for acceptors). Additionally, exonic rare variants that increase the MES donor score of their sequence context and result in a calibrated MES donor score of z>0 had ~64% probability to create a de novo donor, while if they result in a calibrated MES score of -2≤z≤0 the probability to create a de novo donor was ~30% (Spurdle, Goldgar, Parsons, unpublished data). These MES-based rules were used to identify rare sequence variants that are likely to alter mRNA splicing. For exonic variants that resulted in a missense substitution, the higher of the two priors (missense versus splicing) took precedence for multifactorial likelihood analysis. Secondly, revised pathology LRs were drawn from a recent large-scale age-stratified analysis of 4,477 BRCA1 mutation carriers, 2,565 BRCA2 mutation carriers and 47,565 breast cancer cases with no known mutation (Spurdle et al 2015). Additional information included breast tumour ER and grade status extracted from the literature, family history, segregation and co-occurrence likelihood ratios (LRs) assessed from a previously described dataset, and further segregation and pathology data compiled for the variants identified through the kConFab and AFFECT studies. Published mRNA assay data were identified from the literature, where this was available. Current class, and rationale for classification, is summarised in Supp. Table S4 for all test variants.

Authors' contributions
Experiments and analysis were performed by KJF. KJF and NSS participated in sample preparation. M-EB performed pathological review of all tumours. ABS, MP, JM, DEG provided samples and/or participated in data analysis and interpretation. JMF, RB, ABS and JM conceived the study design and coordination. KJF and JMF drafted the manuscript.

**Declarations**

Ethical approval references are listed in the methods. All authors read and approved the final manuscript. No authors have conflicts of interest to declare.

**Acknowledgments**

This study was funded by Breast Cancer Now Project 2010NovPR06. JMF and KJF were funded by Breast Cancer Now. ABS is supported by an NHMRC Senior Research Fellowship. MP was supported by funding from NHMRC project grant ID 1010719. The authors acknowledge funding from Cancer Research UK (A13086) and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. We wish to thank Eveline Niedermayr, all the KConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow up Study (funded by NHMRC grants 145684, 288704 and 454508) for their contributions to this resource, and the many families who contribute to KConFab. KConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council (NHMRC) and by the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. The AFFECT study is funded by Breast Cancer Now and Guy's and St. Thomas' Charity. AFFECT samples contributed by Louise Izatt (Guy’s Hospital, London), Shirley Hodgson (St Georges Hospital, London), Sue Kenwrick (Addenbrooks Hospital, Cambridge) and Fiona Douglas (Freeman Hospital, Newcastle).
References


Figure Legends

**Figure 1: Study Design.** A) Representation of the exonic position of the known mutations in the BRCA1 gene used in this study, including pathogenic (missense and truncating) and test variants. Some variants are represented by multiple tumours. B) Flow diagram illustrating study design. Rectangular boxes indicate samples, hexagonal boxes represent experimental processes, and diamond-shaped boxes denote bioinformatic processes or analysis.

**Figure 2. 450K methylation array analysis defines pathogenic and neutral variants.** A) Venn diagram to represent the comparison of significant probe lists (fdr <0.05), showing 23 probes only significant in the mutation status analysis. B) Consensus clustering using 18 probes with a difference in methylation between BRCA1 mutated and BRCAx greater than 5%, identified two main clusters, as shown by kmeans plot. C) The resulting correlation matrix from these two clusters shows that these clusters correlate with mutation status. Unclassified variants in green clustered with either the BRCA1 tumours or the BRCAx tumours.

**Figure 3. Validation of LYRM9, BACH2, LOC654342 and C8orf31 loci.** A) Strip charts to show array beta values for the four validated loci plotted by mutation status. Blue dotted line is plotted at the median of BRCA1 and BRCAx samples. ER negative samples are coloured red, ER positive samples are coloured blue, and those samples for which ER status is unknown are coloured green (BRCA1 n=18, BRCAx=19, test variant=17, except BACH2 where BRCA1 n=17). B) Strip charts showing the pyrosequencing methylation value of the independent group of samples validating the difference observed on the array (LYRM9: BRCA1 n=19, BRCAx n=29, test variant (UV) n=11. BACH2: BRCA1 n=19, BRCAx n=27, test variant (UV) n=18. C8orf31: BRCA1 n=16, BRCAx n=11, test variant (UV) n=20. LOC654342: BRCA1 n=9, BRCAx n=18, test variant (UV) n=7).

**Figure 4 Combined methylation likelihood ratios.** LRs were calculated for each sample using the model including only methylation. The log of the combined LR is plotted here corresponding to each variant assayed.
Tables

Table 1: Candidate Gene Validation by Pyrosequencing

Table 2: Logistic Regression analysis of Candidate Genes

Table 3: Combined summary predictions of all unknown variants analysed.

Supp. Table S1: Prior candidate gene validation by pyrosequencing

Supp. Table S2: Individual sample predictions for BRCA1 test variant samples analysed.

Supp. Table S3: Individual sample and combined summary predictions for BRCA1 and BRCAx samples analysed.

Supp. Table S4: Current class and rationale for classification of test variants

Supp. Table S5: PCR primers, annealing temperatures, sequencing primer and sequence to analyse used in pyrosequencing assay.
A

- **BACH2**
- **C17orf108**
- **C8orf31**
- **LOC654342**

B

- **BACH2**
- **C17orf108**
- **C8orf31**
- **LOC654342**
### Table 1. Candidate Gene Validation by Pyrosequencing

<table>
<thead>
<tr>
<th>Region Name</th>
<th>Probe ID</th>
<th>wilcox ER q value fdr*</th>
<th>wilcox grade q value fdr#</th>
<th>BACH2 Median Beta Value</th>
<th>BRCA1 Median Beta Value</th>
<th>Methylation difference (beta value)</th>
<th>SNP in probe targeted CpG?</th>
<th>test set</th>
<th>array data (beta values)</th>
<th>Samples Passed QC</th>
<th>Pyrosequencing Validation (% methylation)</th>
<th>ICC: array beta values vs Pyrosequencing %†</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACH2</td>
<td>cg24657115</td>
<td>0.0494</td>
<td>0.0771</td>
<td>0.2378</td>
<td>0.28</td>
<td>0.51</td>
<td>0.23</td>
<td>Y</td>
<td>64 /96</td>
<td>BRCA1 (7/28)</td>
<td>0.0070</td>
<td>35.4</td>
</tr>
<tr>
<td>C8orf31</td>
<td>cg03029255</td>
<td>0.0462</td>
<td>0.1118</td>
<td>0.1711</td>
<td>0.71</td>
<td>0.41</td>
<td>-0.3</td>
<td></td>
<td>47 /24</td>
<td>16</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>C17orf108</td>
<td>cg02502358</td>
<td>0.0494</td>
<td>0.0571</td>
<td>0.4768</td>
<td>0.16</td>
<td>0.07</td>
<td>-0.09</td>
<td></td>
<td>59 /28</td>
<td>19</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>LOCS5A342</td>
<td>cg21645762</td>
<td>0.0494</td>
<td>0.0542</td>
<td>0.2759</td>
<td>0.77</td>
<td>0.61</td>
<td>-0.16</td>
<td></td>
<td>34 /15</td>
<td>9</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>chr13q34</td>
<td>cg12472473</td>
<td>0.0350</td>
<td>0.7011</td>
<td>0.5566</td>
<td>0.91</td>
<td>0.75</td>
<td>-0.16</td>
<td></td>
<td>53 /19</td>
<td>25</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>chr1p34.2</td>
<td>cg07405182</td>
<td>0.0462</td>
<td>0.2613</td>
<td>0.2437</td>
<td>0.87</td>
<td>0.59</td>
<td>-0.28</td>
<td></td>
<td>48 /11</td>
<td>11</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>chr15q22.31</td>
<td>cg08036487</td>
<td>0.0462</td>
<td>0.0542</td>
<td>0.2759</td>
<td>0.75</td>
<td>0.56</td>
<td>-0.19</td>
<td></td>
<td>42 /11</td>
<td>25</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>chr3q25.2</td>
<td>cg02071853</td>
<td>0.0462</td>
<td>0.0633</td>
<td>0.5381</td>
<td>0.72</td>
<td>0.51</td>
<td>-0.21</td>
<td></td>
<td>36 /10</td>
<td>19</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>ZNF319</td>
<td>cg06866605</td>
<td>0.0494</td>
<td>0.0771</td>
<td>0.2696</td>
<td>0.9</td>
<td>0.94</td>
<td>0.04</td>
<td></td>
<td>53 /12</td>
<td>31</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>chr6q27</td>
<td>cg26074411</td>
<td>0.0494</td>
<td>0.1382</td>
<td>0.1243</td>
<td>0.9</td>
<td>0.85</td>
<td>-0.05</td>
<td></td>
<td>33 /6</td>
<td>12</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>INTS1</td>
<td>cg11644627</td>
<td>0.0490</td>
<td>0.0810</td>
<td>0.8266</td>
<td>0.81</td>
<td>0.89</td>
<td>0.08</td>
<td></td>
<td>52 /15</td>
<td>28</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>chr8p23.1</td>
<td>cg04039177</td>
<td>0.0228</td>
<td>0.2284</td>
<td>0.7353</td>
<td>0.57</td>
<td>0.41</td>
<td>-0.16</td>
<td></td>
<td>45 /12</td>
<td>25</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>DLC1</td>
<td>cg00933411</td>
<td>0.0494</td>
<td>0.0849</td>
<td>0.7508</td>
<td>0.37</td>
<td>0.23</td>
<td>-0.14</td>
<td></td>
<td>33 /9</td>
<td>20</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>TJAP1</td>
<td>cg16759204</td>
<td>0.0358</td>
<td>0.1441</td>
<td>0.1885</td>
<td>0.7</td>
<td>0.57</td>
<td>-0.13</td>
<td></td>
<td>51 /13</td>
<td>29</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>PLAG1</td>
<td>cg02697107</td>
<td>0.0494</td>
<td>0.6577</td>
<td>1.0000</td>
<td>0.86</td>
<td>0.81</td>
<td>-0.05</td>
<td></td>
<td>52 /15</td>
<td>28</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>CRTAC1</td>
<td>cg23801028</td>
<td>0.0437</td>
<td>0.1168</td>
<td>0.4680</td>
<td>0.47</td>
<td>0.63</td>
<td>0.16</td>
<td></td>
<td>66 /19</td>
<td>34</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>chr16q24.1</td>
<td>cg03385438</td>
<td>0.0448</td>
<td>0.0602</td>
<td>0.3067</td>
<td>0.9</td>
<td>0.83</td>
<td>-0.07</td>
<td></td>
<td>X</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>KRT6C</td>
<td>cg09440385</td>
<td>0.0494</td>
<td>0.0734</td>
<td>0.5752</td>
<td>0.96</td>
<td>0.91</td>
<td>-0.05</td>
<td></td>
<td>X</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

**Footnotes:**

* False Discovery Rate corrected p value for Wilcoxon rank sum test: BRCA1 mutated tumours vs BRCA1X tumours, beta values, array samples
# False Discovery Rate corrected p value for Wilcoxon rank sum test: ER positive tumours vs ER negative tumours, beta values, array samples
$ False Discovery Rate corrected p value for Wilcoxon rank sum test: grade 1+2 vs grade 3, beta values, array samples
† Intraclass Correlation Coefficient (ICC) calculated using array beta values and pyrosequencing methylation percentages for the same sample (n between 26 and 54, dependent on assay)
ND Not Done - assay design was not possible/could not be optimised
Table 2: Logistic Regression analysis of Candidate Genes

<table>
<thead>
<tr>
<th>Assay</th>
<th>Logistic Regression p values</th>
<th>number of samples included in logistic regression*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mut~meth</td>
<td>mut~meth+ER</td>
</tr>
<tr>
<td></td>
<td>meth</td>
<td>ER</td>
</tr>
<tr>
<td>BACH2</td>
<td>0.0265</td>
<td>0.0607</td>
</tr>
<tr>
<td>C8orf31</td>
<td>0.0003</td>
<td>0.0063</td>
</tr>
<tr>
<td>C17orf108</td>
<td>0.0166</td>
<td>0.1010</td>
</tr>
<tr>
<td>LOC654342</td>
<td>0.0038</td>
<td>0.0301</td>
</tr>
</tbody>
</table>

Footnotes:
* only samples with complete data (methylation, ER status and grade) were used in these logistic regressions
Table 3: Combined summary predictions of all unknown variants analysed.

<table>
<thead>
<tr>
<th>Number of Tumours</th>
<th>Number of Independent Tumours*</th>
<th>HGVS Nuc</th>
<th>HGVS Prot</th>
<th>Prior probability</th>
<th>Current posterior probability†</th>
<th>Current Classification†</th>
<th>Posterior Probability using methylation only</th>
<th>Classification using methylation only</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>c.1486C&gt;T</td>
<td>p.Arg496Cys</td>
<td>0.02</td>
<td>0.00089</td>
<td>Class 1</td>
<td>0.00009</td>
<td>Class 1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>c.2521C&gt;T</td>
<td>p.Arg841Trp</td>
<td>0.02</td>
<td>2.29E-12</td>
<td>Class 1</td>
<td>0.0005-0.0006</td>
<td>Class 1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>c.4963T&gt;C</td>
<td>p.Ser1655Pro</td>
<td>0.03</td>
<td>¥</td>
<td>Class 3</td>
<td>0.39565</td>
<td>Class 3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>c.4103C&gt;T</td>
<td>p.Ala1368Val</td>
<td>0.02</td>
<td>0.00163</td>
<td>Class 2</td>
<td>0.010-0.004</td>
<td>Class 2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>c.823G&gt;A</td>
<td>p.Gly275Ser</td>
<td>0.02</td>
<td>0.00427</td>
<td>Class 2</td>
<td>0.002-0.006</td>
<td>Class 2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>c.5194-12G&gt;A</td>
<td>IVS</td>
<td>0.34</td>
<td>0.99999</td>
<td>Class 5</td>
<td>0.99638</td>
<td>Class 5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>c.4039A&gt;G</td>
<td>p.Arg1347Gly</td>
<td>0.02</td>
<td>2.04E-12</td>
<td>Class 1</td>
<td>0.00967</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.641A&gt;G</td>
<td>p.Asp214Gly</td>
<td>0.02</td>
<td>0.07426</td>
<td>Class 3</td>
<td>0.00065</td>
<td>Class 1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.1984_1992del</td>
<td>p.His662-Arg664del</td>
<td>0.02</td>
<td>0.06058</td>
<td>Class 3</td>
<td>0.04922</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.2912A&gt;G</td>
<td>p.His971Arg</td>
<td>0.02</td>
<td>0.00025</td>
<td>Class 1</td>
<td>0.03682</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.203T&gt;G</td>
<td>p.Ile68Arg</td>
<td>0.66</td>
<td>¥</td>
<td>Class 3</td>
<td>0.01215</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.4185+9C&gt;T</td>
<td>IVS</td>
<td>0.02</td>
<td>0.00232</td>
<td>Class 2</td>
<td>0.00258</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.4484+2_4484+3ins8</td>
<td>IVS</td>
<td>0.97</td>
<td>0.98991</td>
<td>Class 4</td>
<td>0.97881</td>
<td>Class 4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.4485-8C&gt;T</td>
<td>IVS</td>
<td>0.04</td>
<td>¥</td>
<td>Class 3</td>
<td>0.00884</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.5467+5G&gt;C</td>
<td>IVS</td>
<td>0.34</td>
<td>0.19022</td>
<td>Class 3</td>
<td>0.13048</td>
<td>Class 3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.593+16C&gt;G</td>
<td>IVS</td>
<td>0.02</td>
<td>¥</td>
<td>Class 3</td>
<td>0.14272</td>
<td>Class 3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.454C&gt;T</td>
<td>p.Leu152Phe</td>
<td>0.02</td>
<td>¥</td>
<td>Class 3</td>
<td>0.00016</td>
<td>Class 1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.4991T&gt;C</td>
<td>p.Leu1664Pro</td>
<td>0.03</td>
<td>1.18E-05</td>
<td>Class 1</td>
<td>0.22327</td>
<td>Class 3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.1534C&gt;T</td>
<td>p.Leu512Phe</td>
<td>0.02</td>
<td>0.00014</td>
<td>Class 1</td>
<td>0.28730</td>
<td>Class 3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.4955T&gt;A</td>
<td>p.Met1652Lys</td>
<td>0.66</td>
<td>¥</td>
<td>Class 3</td>
<td>0.96673</td>
<td>Class 4</td>
</tr>
<tr>
<td>#</td>
<td>Patient</td>
<td>Variant</td>
<td>Allele</td>
<td>Frequency</td>
<td>LR</td>
<td>Class</td>
<td>Combined LR</td>
<td>Class</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td>-----------</td>
<td>----</td>
<td>-------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.3708T&gt;G</td>
<td>p.Asn1236Lys</td>
<td>0.02</td>
<td>¥</td>
<td>Class 3</td>
<td>0.01645</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.1036C&gt;T</td>
<td>p.Pro346Ser</td>
<td>0.02</td>
<td>0.00427</td>
<td>Class 2</td>
<td>0.00046</td>
<td>Class 1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.2180C&gt;T</td>
<td>p.Pro727Leu</td>
<td>0.02</td>
<td>0.00017</td>
<td>Class 1</td>
<td>0.00352</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.5096G&gt;A</td>
<td>p.Arg1699Gln</td>
<td>0.66</td>
<td>Intermediate risk - based on segregation analysis (Spurdle et al, 2015)</td>
<td>0.97998</td>
<td>Class 4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.5284A&gt;G</td>
<td>p.Arg1762Gly</td>
<td>0.03</td>
<td>¥</td>
<td>Class 3</td>
<td>0.00247</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.551C&gt;T</td>
<td>p.Ser184Phe</td>
<td>0.02</td>
<td>¥</td>
<td>Class 3</td>
<td>0.01149</td>
<td>Class 2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c.1423A&gt;T</td>
<td>p.Ser475Cys</td>
<td>0.02</td>
<td>0.00395</td>
<td>Class 2</td>
<td>0.02111</td>
<td>Class 2</td>
</tr>
</tbody>
</table>

Footnotes:
* = only tumours from different individuals can be considered independent
† = based on segregation analysis alone (Arg1699Gln) or multifactorial likelihood analysis incorporating segregation, pathology, and other data points (see supplementary table 2)
¥ = Combined LR does not pass thresholds recommended as per ENIGMA BRCA classification guidelines (http://www.enigmacconsortium.org/), namely LR of <0.5 (to reach final class 2 or 1), or >2.0 (to reach final class 4 or 5). and so should be considered class 3 uncertain.