Whole exome DNA sequence analysis of *Brca2* and *Trp53* deficient mouse mammary gland tumours

Jeffrey C. Francis[1,2], Lorenzo Melchor[1,4], James Campbell[1,2], Howard Kendrick[1,5], Wenbin Wei[1,2], Javier Armisen-Garrido[3], Ioannis Assiotis[3], Lina Chen[3], Iwanka Kozarewa[3], Kerry Fenwick[3], Amanda Swain[3], *Matthew J. Smalley[1,5], *Christopher J. Lord[1,2] and *Alan Ashworth[1,2]

[1]The Breakthrough Breast Cancer Research Centre, [2]CRUK Gene Function Laboratory and [3]Tumour Profiling Unit, The Institute of Cancer Research, London, SW3 6JB, UK.

[4]Current address: Division of Molecular Pathology, The Institute of Cancer Research, London, SM2 5NG, UK.

[5]Current address: European Cancer Stem Cell Research Institute, University of Cardiff, Cardiff, CF24 4HQ, UK.

*Email: SmalleyMJ@cardiff.ac.uk, Chris.Lord@icr.ac.uk, Alan.Ashworth@icr.ac.uk

Running title: DNA sequencing of *Brca2* and *Trp53* deficient mammary tumours

Competing interests: The authors declare that they have no competing interests.

Abstract

Germline mutations in the tumour suppressor *BRCA2* predispose to breast, ovarian and a number of other human cancers. *Brca2* deficient mouse models are used for pre-clinical studies but the pattern of genomic alterations in these tumours has not yet been described in detail. We have performed whole exome DNA sequencing analysis of mouse mammary tumours from *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ animals, a model of *BRCA2* deficient human cancer. We also used the sequencing data to estimate DNA copy number alterations in these tumours and identified a recurrent copy number gain in *Met*, which has been found amplified in other mouse mammary cancer models. Through a comparative genomic analysis, we identified several mouse *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mammary tumour somatic mutations in genes that are also mutated in human cancer, but few of these genes have been found frequently mutated in human breast cancer. A more detailed analysis of these somatic mutations revealed a set of genes that are mutated in human *BRCA2* mutant breast and ovarian tumours that are also mutated in mouse *Brca2* null, *Trp53* null mammary tumours. Finally, a DNA deletion surrounded by microhomology signature found in human *BRCA1/2* deficient cancers was not common in the genome of these mouse tumours. Although a useful model, there are some differences in the genomic landscape of tumours arising in *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mice compared to human *BRCA2* mutated breast cancers. Therefore this needs to be taken into account in the use of this model.

Keywords: *Brca2*, mammary gland, mouse model, exome sequencing

Introduction

Women with germline loss of function mutations in the *BRCA2* gene have a significantly increased lifetime risk of developing breast and ovarian cancers [1]. Mutations in *BRCA2* also predispose to developing cancers of the pancreas, stomach, gallbladder, bile duct, skin, and in men, the prostate [2]. Germline *BRCA2* mutations are found in approximately 3% of breast and 8% of ovarian cancers, whilst *BRCA2* is somatically mutated in 2% of breast and 3% of ovarian cancers [3, 4].

BRCA2 functions as a tumour suppressor, and somatic loss of the second, wild type allele in tumours is a common event. *BRCA2* encodes a protein that has a key role in the accurate repair of DNA double-strand breaks (DSBs) through the process of homologous recombination (HR) [5]. When HR is active, the homologous DNA sequence to the DNA flanking the DSB is used as a template that guides DNA synthesis on the damaged sister chromatid, resulting in high-fidelity repair. During this process, BRCA2 interacts with and ensures the loading of the DNA recombinase RAD51 onto DNA at the site of the DSB. The formation of RAD51/DNA nucleoprotein complexes facilitates strand invasion, a critical step in the repair of DSBs by HR. When BRCA2 is defective, DSBs fail to be repaired by HR and more error-prone mechanisms of DNA repair are used [6, 7]. In some cases, the enhanced utilisation of error-prone mechanisms of DNA repair can lead to genomic instability and ultimately tumour development [reviewed in [8]]. Consistent with its role in HR, *BRCA2* deficient human tumours tend to be associated with mutations in the *TP53* gene, with mutations identified in up to 60% of *BRCA2* mutant breast cancers [9] and 70% of *BRCA2* mutant ovarian cancers [10]. *BRCA2* mutant human tumours are characterised by a particular mutational signature, namely base substitutions and an enrichment of large deletions (up to 50 bp) which have breakpoints that are flanked by regions of DNA sequence microhomology [11-13]. These latter mutations are

possibly the result of error-prone homology directed DNA repair in tumour cells that have defective HR [6, 7, 14]. The HR defect in *BRCA2* mutant tumours (and also *BRCA1* mutant tumours) has recently been exploited in the design of a synthetic lethal treatment approach that exploits small molecule inhibitors of the DNA repair proteins Poly-(ADP-ribose) Polymerase (PARP)-1/2 (reviewed in [15]). PARP1 is an enzyme involved in the repair of single-strand DNA breaks and inhibitors of this protein result in the stalling of replication forks. HR repairs the DNA lesions induced by PARP1 inhibitors in wild type cells, making HR defective *BRCA2* mutant cells highly sensitive to these therapeutics.

Genetically engineered mouse models (GEMM) of human cancer have been used as a tool to study genetic lesions thought to promote tumour development. Many of these models recapitulate the molecular and cellular features of human cancer and have provided insights into the processes that drive tumorigenesis [16, 17]. GEMMs have also allowed investigators to test the response of novel cancer therapeutics in pre-clinical *in vivo* models that have predefined genetic aberrations [18]. For example, *Brca2* mutant mice with mammary tumours show an anti-tumour response to PARP1/2 inhibitors [19, 20]. In addition, DNA sequence analyses of tumours from GEMMs have identified secondary mutations in cancer-promoting genes that are also mutated in human tumours [21-24]. These comparative genomics approaches have shown that genes mutated in both species can promote tumour formation and provide a method for identifying cancer targets within the setting of defined genetic initiating event.

In this study we used whole exome DNA sequencing of *Brca2* null, *Trp53* null mammary tumours from mice to identify additional somatic mutations associated with these cancers. We compared the somatic mutations we identified in this mouse

model with the genes mutated in human cancer and the mutations found in human *BRCA2* mutant breast and ovarian cancers.

Materials and Methods

Generation of *Brca2* null, *Trp53* null mammary tumours

Mice carrying the *Blg-Cre* transgene [25] were mated with animals with both *loxP* flanked *Brca2* and *loxP* flanked *Trp53* containing alleles [26] generating *Blg-Cre* $Brca2^{f/f}$ $Trp53^{f/f}$ animals [27]. Mammary tumour and spleen tissue from mutant animals were excised from humanely killed tumour bearing mice. Part of the tumour was fixed in 4% phosphate-buffered formalin overnight for histological analysis and immunohistochemistry and part was snap frozen on dry ice for isolation of DNA.

Genomic DNA preparation

Genomic DNA was isolated from tumours and spleens using the DNeasy Blood and Tissue kit (Qiagen, UK), according to the manufacturer's instructions. DNA was quantified using the Qubit 2.0 kit (Life Technologies, UK).

Whole exome DNA sequencing

Three micrograms of genomic DNA was fragmented to 200 bp using a Covaris E Series instrument (Covaris Inc, MA, USA) and the resultant library subjected to DNA capture using the 50 Mb SureSelect Mouse All Exon kit (Agilent, CA, USA). DNA capture was carried out according to the manufacturer's instructions. Illumina paired-end libraries were prepared from the captured target regions and quantified using a Bioanalyzer DNA chip (Agilent), followed by sequencing on a HiSeq2000 platform (Illumina, San Diego, CA, USA), acquiring 2 × 76 bp reads. Casava software (v1.8, Illumina) was used to make base calls. Sequences were output in fastq format. Reads failing the Illumina chastity filter were removed before further analysis. The

raw data from this sequencing procedure is now deposited on the European Nucleotide Archive [ENA Accession Number: PRJEB6674; http://www.ebi.ac.uk/ena/data/view/PRJEB6674].

Read mapping and detection of somatic mutations in exome sequencing

Burrows-Wheeler Alignment (BWA, v0.5.9) was used to align reads to the mouse reference genome (GRCm37) [28]. Duplicate sequence reads (PCR-derived duplicates) were removed from further analysis at this point. Base quality recalibration, realignment around indels, and variant calling were performed using the Genome Analysis Tool Kit (GATK, v1.0-6144-gdd92a14) using the Broad best practice variant detection workflow [29]. The MuTect algorithm (v1.1.4) was also used to identify somatic single nucleotide mutations in targeted exons [30]. Small insertions and deletions detected in the tumour sample that were absent in the matched normal were considered to be candidate somatic mutations. Variants called in regions not covered by the exome capture probes and variants marked as low quality (QUAL below 20) were excluded. Candidate somatic mutations were also assessed to confirm their validity by visualizing sequencing data using the Broad Integrative Genomics Viewer tool [31]. The PROVEAN (Protein Variation Effect Analyzer, v1.1.3) software tool was used to predict whether a mutation has an impact on the biological function of a protein.

Detection of copy number alterations in exome sequencing

Copy number alterations were predicted using exome DNA sequencing data and the CONTRA (v2.0.4) and CoNIFER (v0.2.2) software packages using default parameters [32, 33].

Validation sequencing

To validate somatic mutations by Sanger sequencing, PCR amplicons encompassing the candidate mutation sites were sequenced in tumour and spleen DNA using standard methods. PCR primers for amplification and sequencing were designed using the UCSC *In Silico* PCR tool [34].

Comparison with human somatic mutation data

The likely human orthologs of the mouse *Brca2* null, *Trp53* null deficient mammary tumour somatic mutations were identified using the MGI curated sets of homology (from NCBI HomoloGene build67) [35]. Tumour DNA sequence data from non-familial human breast cancer and ovarian cancer was accessed via the TCGA data portal [36].

Fluorescence In Situ Hybridisation (FISH)

BAC clones mapping to the *Met* locus (6A2) were used as FISH probes: RP23-239E3 (mid-position 17462960), RP23-73G15 (mid-position 17571929), RP23-444N4 (mid-position 17647861). Additionally, three BAC clones from the 6G3 chromosomal region (close to the telomere) were used as copy number reference for mouse chromosome 6: RP23-27F7 (mid-position 147281394), RP24-127P8 (mid-position 147421860), RP23-396J18 (mid-position 147597428). BACs were ordered from BACPAC Resources Center at the Children's Hospital, Oakland Research Institute (CHORI, Oakland, CA; http://bacpac.chori.org). These were labelled by nick translation with SpectrumOrange-dUTP (for *Met* probes) and SpectrumGreen-dUTP (for reference probes) (Vysis Inc, Downers Grove). Routine FISH protocol for formalin-fixed paraffin embedded tissues was used with slight modifications (described in Supporting Methods).

Results

Exome sequencing of *Brca2* deficient mouse tumours

We wanted to study the pattern of genomic alterations present in *Brca2* deficient tumours, and in particular, we sought to determine if functional disruption of the *Brca2* gene would lead to recurrent somatic mutations in a mouse model of mammary cancer. To test this, we carried out whole exome DNA sequencing of *Brca2* null, *Trp53* null mammary tumours arising in virgin and parous *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ conditional knock-out mice. To generate these tumours, mice carrying the *Blg-Cre* transgene [25] were mated with animals with both *loxP* flanked *Brca2* and *loxP* flanked *Trp53* containing alleles designed to delete *Brca2* and *Trp53* in the epithelia of the mammary gland upon Cre-mediated recombination [26]. These tumours are responsive to PARP1/2 inhibitors [19], drugs that target *Brca2* deficient cells [37, 38].

The median mammary tumour latency in virgin mice was 242 days and in parous animals was 275.5 days. [27]. Eight of the tumours in this cohort (904, 907, 918, 919, 1001, 1008, 1015 and 1016) have previously been described [27]; their key features along with four additional tumours (1034, 1035, 1103 and 1114) are described in Supporting Table 1. Altogether, this cohort of tumours consisted of seven tumours diagnosed as Invasive Ductal Carcinoma of No Special Type (IDC-NST), four Metaplastic Spindle Cell Tumours (MSCTs), and one Adenomyoepithelioma (AME). These tumours were high grade, with a high mitotic index and exhibited nuclear polymorphism (Table S1 and [27]). Only 2/11 (18%) were estrogen receptor (ER) positive (cut-off for ER positivity >5% ER expressing cells).

DNA from *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mammary tumours was subjected to exome capture and sequencing using a Illumina HiSeq 2000 platform. A schematic of the workflow is shown in Figure S1. We also exome captured and sequenced DNA

derived from the normal spleen of each animal. A total of 12 tumours and matched spleens were sequenced from 12 *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* animals. An average of 75 million sequencing reads was generated for each sample giving a median depth of 42x (Table S2). In these mouse tumour samples the most recurrently mutated genes in human breast cancer were each DNA exome sequenced at > 10x (Table S3). In this mouse model, Cre-mediated deletion is expected to result in the loss of the *loxP* flanked *Brca2* exon 11 and the *loxP* flanked *Trp53* exons 2-10. Visualization of the DNA exome sequencing reads and analysis of the RPKM (reads per kilobase per million) values confirmed that in each tumour there was a clear loss of *Brca2* exon 11 and *Trp53* exons 2-10 when compared to the matched spleens (Figure 1 and Table S4). The relative frequency of *Brca2* and *Trp53* exons not flanked by *loxP* sites remained unchanged in all tumour/spleen comparisons (Figure 1 and Table S4). This data is in agreement with our previous droplet digital PCR data that demonstrated that these tumours had fewer copies of the *loxP* flanked *Brca2* and *Trp53* exons compared to the exons outside of these regions [27]. Histologically, all samples were at least 80% tumour cells [27].

Somatic mutations and copy number alterations in mouse *Brca2* null, *Trp53* null mammary tumours

To identify somatically occurring mutations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours, we compared exome sequence data from tumour DNA and matched spleen DNA. From 12 tumour/spleen comparisons, we identified 963 candidate somatic mutations in a total of 657 genes (Table S5 and Table S6). There were between 41 and 129 candidate alterations in each tumour sample, a range which is comparable to the average of 60 exomic mutations identified in human breast tumours [4] (Figure 2A and Table S7). 714 of these candidate mutations (74%) were base substitutions, 144 (15%) were insertions and 105 (11%) deletions (Figure 2B and Figure 2C). The

seven IDC-NST tumours had an average of 81 candidate somatic mutations, the four MSCTs had an average of 73 somatic mutations and the one AME tumour had 101 somatic mutations (Table S8). Of the 963 candidate somatic mutations, 290 were predicted to cause an amino acid sequence alteration in a total of 241 genes (Table S9), with 22 (8%) being potential frameshift mutations, 13 (4%) nonsense mutations, 225 (78%) missense mutations and 30 (10%) splice site mutations (Figure 2D, Figure 2E and Table S7). Of the 241 genes predicted to have coding mutations in the mouse mammary gland tumours, 191 likely human orthologs were identified using the Mouse Genome Informatics curated sets of homology (Table S9). To be confident in our exome sequence variant calling, we also assessed 14 randomly selected candidate somatic mutations using Sanger sequencing. The majority of these mutations (17/23) were confirmed by this orthogonal sequencing method, giving us some confidence in our variant detection approach (Table S10).

We also used the exome DNA sequencing data to estimate the presence of somatically occurring copy number alterations in mammary gland tumours from *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mice. To do this analysis we used two different software packages, CONTRA (COpy Number Targeted Resequencing Analysis) and CoNIFER (Copy Number Inference From Exome Reads) [32, 33]. Using the exome data from the 12 *Brca2* null, *Trp53* null tumours, CONTRA identified a total of 1149 candidate target region gains and 338 candidate target region losses (Figure 3A), with between 1 and 413 gains and 4 and 141 losses identified in each tumour sample (Figure 3B and Figure 3E) (Table S11). CoNIFER identified a total of 146 candidate copy number gains and 45 candidate copy number losses (Figure 3C), with between 2 and 24 gains and 0 and 10 losses identified in each tumour sample (Figure 3D and Figure 3E) (Table S12). Analysis of the data output from CONTRA and CoNIFER revealed a total of 87 genes predicted to have a DNA copy number alteration by both packages (Table S13).

Recurrent somatic mutations and copy number alterations

We searched for recurrently mutated genes amongst the 12 *Brca2* null, *Trp53* null mammary tumours. This analysis identified three genes that appeared to be mutated in at least two tumours which had clear human orthologs : *Sik1* (*Salt inducible kinase 1*), *Prr14l* (*Proline rich 14-like*) and *Cd244* (*CD244 natural killer cell receptor 2B4*).

Two *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ tumours, arising in different mice, had somatic mutations in *Sik1*, a gene that encodes a serine/threonine-protein kinase. These were frameshift mutations that caused a p.L741fs alteration located at the C-terminal end of the protein. We did note however that this mutation was present within a homopolymer stretch and thus could be an artifact of the deep sequencing approach used. We used Sanger sequencing to assess this possibility and were not able to validate the presence of *Sik1* mutations in either original tumour (data not shown). We identified three mammary tumours with different missense mutations in the receptor domain of *Cd244* (p.H61D, p.D72A [this amino acid is not conserved in the human protein] and p.S80F), a cell surface receptor expressed on natural killer cells and T cells [39]. Each of these missense mutations was clustered within the first Ig-like 1 domain (Figure 4A). We also noted two tumours with mutations in *Prr14l*, both being frameshift mutations, p.E236fs and p.D451fs, however only the latter mutation validated (Table S10). *CD244* and *PRR14L* are both somatically mutated in human cancer, with *CD244* mutations present in bladder cancer (3.6% of TCGA samples) and melanoma (3.6% of TCGA samples), while *PRR14L* is mutated in uterine cancer (6% of TCGA samples) [40]. At present, few mutations have been identified in either gene in human breast cancer (*CD244* 0.2% of TCGA samples; *PRR14L* 0.4% of TCGA samples).

We also searched for recurrent somatically occurring DNA copy number alterations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* tumour samples. Using the data obtained from CONTRA and CoNIFER we identified a candidate recurrent gain on chromosome 6 in 5/12 (42%) *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours we sequenced (Figure 4B). Analysis of the DNA sequence from these samples using IGV demonstrated there was an increase in tumour DNA sequence reads in this region of chromosome 6 compared to the matched spleen (Figure S2). The smallest amplicon on chromosome 6 in these tumours included the genes *Tcfec*, *Tes*, *Cav2*, *Cav1*, *Met*, *Capza2* and *St7*. We used fluorescent in situ hybridisation (FISH) analysis to confirm this amplification in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. Of the five tumours predicted by CONTRA and CoNIFER to have a chromosome 6 gain all five showed a *Met* FISH probe fluorescent signal consistent with *Met* amplification (Figure 4C). *Met*, a tyrosine kinase receptor for hepatocyte growth factor, has been shown to induce mammary tumours with diverse histologies, including metaplastic features, that are associated with aggressive human basal-like breast cancers and *Met* can synergise with *Trp53* loss to promote mammary tumours with a claudin-low morphological phenotype [41-43]. Copy number gains in the chromosomal region that contains *Met* have also been found in a number of other mouse mammary tumour models [44-48]. The human orthologs of the genes amplified in the chromosome 6 region in the mammary tumours (*TFEC*, *TES*, *CAV2*, *CAV1*, *MET*, *CAPZA2* and *ST7)* have a copy number gain in 1% (8/1062) of breast tumours and 1.6% (5/316) of ovarian cancers, suggesting this is not a common event in the human disease [3, 4]. None of the human tumours that have this amplification event have a loss of function mutation in *BRCA2*.

Comparison of *Brca2* null, *Trp53* null mammary tumour somatic mutations with human cancer

We also assessed whether the non-recurrent coding mutations identified in the *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/}$* mammary tumours were present in genes likely to be drivers of human cancer development. To do this, we cross referenced the list of 191 genes with coding mutations in the *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/}$* mammary tumours (those with likely human orthologs) with datasets describing the likely driver mutant genes in human cancer [40, 49]. This analysis identified six genes that are mutated in this mouse model that are also significantly mutated in human cancer, which is a greater overlap of mutated genes than expected by chance (two-tailed Fisher Exact test p=0.016, odds ratio 3.1 (95% CI 1.1-7.0)) (Table 1 and Figure 4A). All of these mammary tumour mutations validated using Sanger sequencing (Table S10). Two of these genes are frequently mutated in human breast cancer: *GATA3* (10.6% of TCGA samples) and *NCOR1* (4% of TCGA samples). Both of these genes are predominantly mutated in the luminal A subtype of the disease and in general exhibit loss of function mutations [4, 50]. The *Gata3* mutation we identified in the mammary tumour, p.D6E, is unlikely to be a driver mutation as the majority of human breast cancer associated mutations in *GATA3* are splice site or frameshift mutations that encode proteins with truncations from amino acid residue 300 onwards. The Protein Variation Effect Analyzer (PROVEAN) tool predicted *Gata3* p.D6E to have a neutral effect on protein function. In addition, in contrast to the majority of human breast tumours, this *Gata3* mutation was found in an ER negative mammary tumour. One *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour had a *Ncor1* missense mutation, p.V227A, that is within a region of the protein that mediates the transcriptional repressor function of NCOR1, suggesting some functional significance [51, 52]. Conservation of the human and mouse NCOR1 protein sequences in this region (96% identity) also suggest that this domain might be functionally important, although the p.V227 residue is not conserved between species. Four other genes implicated as likely drivers of human cancer were also identified in this comparative analysis, although their frequency of mutation in human breast cancer is relatively low (0.3 % - 1.1 % of

TCGA breast cancers). These included; (i) a missense mutation in *Kras* encoding an alteration at amino acid 13 (p.G13R), (ii) a mutation in a *Notch2* gene splice site (c.2601+6G>A), (iii) a deletion in *Bcor* (p.P1099_1110Ydel) and (iv) a missense mutation in a leucine-rich repeat of *Tlr4* (p.A132del) (Figure 4A and Figure 4D). *KRAS* (*v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*) is frequently mutated in human cancer (57.9% TCGA pancreatic adenocarcinoma; 43% TCGA colorectal adenocarcinoma; 32.6% TCGA lung adenocarcinoma) but is only found altered at a low frequency in human breast cancer (0.8% TCGA breast cancer). Mutations at *KRAS* amino acid 13 are likely cancer drivers, and encode a constitutively active form of the protein [53]. As expected, PROVEAN predicted that the p.G13R mutation altered *Kras* protein function.

We also compared our data set to the mutational profile of human tumours with somatic *BRCA2* mutations. To do this we used mutation data from the TCGA, including exome sequence data from 778 invasive breast carcinoma samples and 520 ovarian serous cystadenocarcinoma samples [3, 4]. These data included 13 *BRCA2* mutated breast cancers (Table S14 and Table S15) and 13 *BRCA2* mutated ovarian cancers (Table S16 and Table S17). We also included mutation data from five additional *BRCA2* mutated breast samples sequenced as part of the ICGC consortium [54, 55] (Table S14 and Table S15).

We compared the somatic mutations we identified in *Brca2* null, *Trp53* null mouse mammary tumours with the genes mutated in human *BRCA2*-associated breast cancers. This comparison identified 31 genes mutated in both cancer types (Figure 5A and Table S18). These genes included *GATA3* and *CD244*, a recurrently mutated gene in the *Brca2* null mouse model. Of these 31 genes, three genes are recurrently mutated in human *BRCA2* mutant breast cancer and included genes thought to be involved in cancer development. For example, *RYR2* is a member of the ryanodine

receptor family, which encode proteins that form calcium channels and have been implicated in breast cancer risk and tumour grade [56, 57]. *SPEN* (*split ends, homolog of Drosophila*) an RNA-binding coregulatory protein and negative regulator of the Notch pathway was identified as a cancer associated gene in adenoid cystic carcinoma [58], and *SYNE1* encodes a spectrin repeat containing protein that localizes to the nuclear envelope and has been reported to be mutated in several cancer types [1, 2, 5]. There were 11 genes in common between the mammary tumour mutations and genes mutated in human *BRCA2* mutant, *TP53* mutant breast tumours (Figure 5B and Table S18). A similar analysis involving a comparison of the *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ tumour mutations with the genes mutated in human *BRCA2* mutant ovarian cancers identified 16 genes, while 10 genes were also mutated in *BRCA2* mutant, *TP53* mutant ovarian cancers (Figure 5C, Figure 5D and Table S19). These included *RYR2* and *SYNE1*, as well as two human tumours with mutations in *VPS13B*, a gene that encodes a member of the vacuolar protein-sorting pathway previously implicated in tumorigenesis [59-61]. Of the genes mutated in both human and mouse *Brca2/BRCA2* mutant tumours, *RYR2*, *SYNE1*, *SPEN* and *VPS13B* are recurrently mutated in the human disease. None of the mutations we identified in these mouse genes have reported mutations at the same position in human cancer. Furthermore, *RYR2*, *SYNE1*, *SPEN* and *VPS13B* are all large genes (> 12 kb coding sequence), suggesting that the mutations in these genes might be due to their size rather than being selected for as driver events.

Base substitutions and microhomology

Recent whole genome sequencing studies have uncovered the patterns of mutational signatures present in human breast cancer [12]. Detailed analysis of the spectrum of mutation types revealed that tumours could be subgrouped based on these alterations, which are likely due to common exposure to mutagens or particular

underlying DNA damage/repair mechanisms [13]. Breast tumours with either a *BRCA1* or *BRCA2* mutation display a mutation signature that clusters these cancers together, namely a relative increase in deletions at regions of DNA sequence flanked by microhomology, thought to be the result of non-homologous end joining (NHEJ) double-strand break repair, and a base substitution signature [12, 13, 62]. To investigate whether the mutational prolife of mouse *Brca2* null, *Trp53* null mammary tumours was similar to human *BRCA* mutant breast cancer, we compared the pattern of exome mutations from *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary cancers to the signature of mutations in human *BRCA* mutant breast cancers. Analysis of the proportion of base substitutions revealed that, similar to human breast tumours, the predominant base substitution in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours was C>T (Figure 6A), which can arise from deamination of methylated cytosine to thymine [63]. Using the previously described computational framework to characterise mutational signatures [64], we were able to derive one reproducible signature (stability > 0.96) from the *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour DNA base substitution exome data. This signature accounts for 96% (688/714) of the substitution mutations and is characterised by C>T substitutions (Figure 6B and Figure 6C). This analysis indicates that the mouse *Brca2* null, *Trp53* null mammary tumour mutational signature exhibits the substitution of some bases more than other bases (for example C>T), compared to the relatively even distribution of substitutions observed in the human *BRCA1/2* mutant signature [12, 13].

We also analysed *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* exome data for the presence of deletions flanked by regions of DNA sequence microhomology. On average, each tumour exhibited 8 deletion events in the exomic sequence, allowing the analysis of 105 deletion events in total (Figure 7A). Of these the majority were 1 bp deletions (Figure 7B and Figure 7C). However, in contrast to human *BRCA* mutant tumours and also mouse *Brca1* mutant tumours [12, 21], the majority of deletions in *Blg-Cre Brca2$^{f/f}$*

Trp53$^{f/f}$ mammary tumours did not exhibit flanking sequence microhomology (Figure 7D).

Discussion

Using next generation DNA sequencing, we have studied the exomic mutations present in mammary tumours from *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mice. As this and similar models provide a test bed for assessing therapeutic approaches to cancer and have also been used to dissect the natural history and pathology of the disease, an understanding of their genetic and genomic composition is important.

In many respects, mammary tumours in *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mice resemble the equivalent human breast cancers. This is perhaps best exemplified by the response of tumours in this model to therapeutic approaches such as PARP inhibitors and platinum salts that target the defect in HR caused by loss of Brca2 function [19, 20]. However, in terms of the types of mutation found in mammary tumours in *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mice, there is seemingly less concordance between this model and the mutational spectrum of *BRCA2* mutant breast tumours, an observation consistent with previous low depth whole genome DNA sequencing efforts [21]. For example, few of the somatically occurring mutations identified in this study have, as yet, been highlighted as candidate driver mutations in human breast tumours. In addition, the characteristic mutational pattern seen in human *BRCA2* mutant cancers, a relatively high frequency of DNA deletions flanked by regions of microhomology, is seemingly uncommon in the mouse model studied here, despite this being observed in *Brca1* mutant mice [21]. Although the reasons for these differences are not as yet clear, there are a number of possible explanations. Firstly, although the mutations we identified in the *Blg-Cre Brca2*$^{f/f}$ *Trp53*$^{f/f}$ mammary tumours were relatively rare in human breast cancer, we note that large datasets focusing exclusively on tumour-

specific mutations in familial *BRCA2* mutant breast cancer, as opposed to sporadically occurring disease, are not as yet available. It is therefore possible that a concordance between the mutational spectrum of tumours in the mouse model studied here and the human disease might not be apparent until larger numbers of familial *BRCA2* mutant tumours have been sequenced. Secondly, the distinct mutational spectrum in our mouse model compared to the human disease might be reflective of differences in etiology. For example, in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* tumours the timing and order in which *Brca2* and *Trp53* alleles are deleted in the mammary gland might not fully reflect pathogenesis in human *BRCA2* mutant cancers and this might influence the mutational spectrum of mammary tumours in the mouse model. Furthermore, the latency of *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours is relatively short (242 - 275.5 days) compared to human tumour formation. This relatively short latency could result in less time to accumulate somatic mutations and genomic signatures that are present in human breast cancer. It will be interesting in the future to test whether the serial transplantation and subsequent outgrowth of *Brca2* deficient mammary tumours might lead to an increase in genomic instability that more closely resembles human breast cancer. Finally, the cell of origin in which the initiating genetic lesion has occurred could also affect the spectrum of mutations. In this study, the *Blg-Cre* transgene deleted *Brca2* and *Trp53* preferentially in luminal ER negative mammary epithelial cells, resulting in tumours that were predominantly ER negative and have a basal-like profile [27]. The majority of *BRCA2* deficient human breast tumours are ER positive and are not basal-like. These issues however do not necessarily negate the utility of such models in assessing some therapeutic strategies, for example PARP inhibitions, which are showing promise as a therapy in human *BRCA* mutant tumours and have anti-tumour efficacy in mouse *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours [19, 65, 66]. Nevertheless, some caution should be used when directly inferring the mechanisms of resistance to therapies as the different mutational spectrum present in this model could affect the mode of

resistance. Despite the apparent differences in the mutational spectrum of the disease, elements of the underlying biology of mouse mammary tumours might accurately reflect the human disease. For example, even though the mutational spectrum might seem distinct, the biological processes that are driven by mutant genes in the mouse model might also operate in the human disease. It would be interesting to test whether mutations identified in this study do indeed accelerate tumour formation in a *Brca2* deficient background. If this were the case, this would strengthen the argument for assessing whether the pathways that these genes modulate are also altered in human *BRCA2* mutant breast cancers.

Acknowledgements

This work was funded by Cancer Research UK and Breakthrough Breast Cancer. We thank Olivia Fletcher for her helpful discussions on the statistical analysis, and to Manisha Maurya and Maria José Carnicer for assistance with the FISH analysis.

Authors' contributions

Conceived and designed the experiments: JCF, LM, MJS, CJL and AA. Performed the experiments: JCF, LM, JC, WW, HK, JAG, IA, LC, IK, KF and AS. Analysed the data: JCF, JC, CJL and AA. Wrote the manuscript: JCF, MJS, CJL and AA. All authors read and approved the final manuscript.

Supporting information

Table S1. Histopathalogical features of *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours used in this study.

Table S2. Quantity and coverage of DNA exome sequencing data.

Table S3. Median depth of sequence in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours of the most recurrently mutated genes in human breast cancers.

Table S4. Log2(RPKM+1) values for *Brca2* and *Trp53* sequence.

Table S5. Unfiltered somatic mutations identified in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours.

Table S6. Filtered somatic mutations identified in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours.

Table S7. Number and types of somatic mutations detected in each mouse mammary tumour.

Table S8. Number and types of de novo mutations detected in each histological type.

Table S9. Somatic mutations predicted to alter protein sequence in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours.

Table S10. Sanger sequenced mutations

Table S11. CONTRA identified DNA target region alterations.

Table S12. CoNIFER identified DNA copy number alterations.

Table S13. Genes with copy number alteration detected by CONTRA and CoNIFER.

Table S14. TCGA and ICGC BRCA2 mutant breast tumours

Table S15. TCGA and ICGC *BRCA2* mutant breast cancer somatic mutations.

Table S16. TCGA BRCA2 mutant ovarian tumours

Table S17. TCGA *BRCA2* mutant ovarian cancer somatic mutations.

Table S18. Comparison of genes mutated in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours and human *BRCA2* mutant breast tumours.

Table S19. Comparison of genes mutated in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours and human *BRCA2* mutant ovarian tumours.

Figure S1. Schematic of the workflow to exon sequence *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. Spleen and mammary tumour was dissected from mutant *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* animals and genomic DNA was isolated. DNA exon sequencing was performed using Illumina technology, followed by BWA alignment, variant calling and somatic tumour mutation determination.

Figure S2. *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours have a candidate gain on chromosome 6. Left panels, *Met* exons 3-7 DNA sequence reads displayed in the Broad Institute IGV show an increase in tumour reads compared to the matched spleen. Right panels, the tumour DNA sequence read copy number gain break point indicated with a black arrow. Spleen and tumour reads from each animal are indicated.

References

1. Antoniou A, Pharoah PD, Narod S*, et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003; **72**: 1117-1130.
2. Jemal A, Siegel R, Ward E*, et al.* Cancer statistics, 2008. *CA Cancer J Clin* 2008; **58**: 71-96.
3. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011; **474**: 609-615.
4. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490**: 61-70.
5. Gudmundsdottir K, Ashworth A. The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene* 2006; **25**: 5864-5874.
6. Moynahan ME, Pierce AJ, Jasin M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol Cell* 2001; **7**: 263-272.
7. Tutt A, Bertwistle D, Valentine J*, et al.* Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. *EMBO J* 2001; **20**: 4704-4716.
8. Venkitaraman AR. Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. *Science* 2014; **343**: 1470-1475.
9. Crook T, Brooks LA, Crossland S*, et al.* p53 mutation with frequent novel condons but not a mutator phenotype in BRCA1- and BRCA2-associated breast tumours. *Oncogene* 1998; **17**: 1681-1689.

10. Rhei E, Bogomolniy F, Federici MG*, et al.* Molecular genetic characterization of BRCA1- and BRCA2-linked hereditary ovarian cancers. *Cancer Res* 1998; **58**: 3193-3196.

11. Barber LJ, Rosa Rosa JM, Kozarewa I*, et al.* Comprehensive genomic analysis of a BRCA2 deficient human pancreatic cancer. *PLoS One* 2011; **6**: e21639.

12. Nik-Zainal S, Alexandrov LB, Wedge DC*, et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; **149**: 979-993.

13. Alexandrov LB, Nik-Zainal S, Wedge DC*, et al.* Signatures of mutational processes in human cancer. *Nature* 2013; **500**: 415-421.

14. Moynahan ME, Chiu JW, Koller BH*, et al.* Brca1 controls homology-directed DNA repair. *Mol Cell* 1999; **4**: 511-518.

15. Lord CJ, Ashworth A. Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nat Med* 2013; **19**: 1381-1388.

16. Tuveson DA, Jacks T. Technologically advanced cancer modeling in mice. *Curr Opin Genet Dev* 2002; **12**: 105-110.

17. Heyer J, Kwong LN, Lowe SW*, et al.* Non-germline genetically engineered mouse models for translational cancer research. *Nat Rev Cancer* 2010; **10**: 470-480.

18. Sharpless NE, Depinho RA. The mighty mouse: genetically engineered mouse models in cancer drug development. *Nat Rev Drug Discov* 2006; **5**: 741-754.

19. Evers B, Drost R, Schut E*, et al.* Selective inhibition of BRCA2-deficient mammary tumor cell growth by AZD2281 and cisplatin. *Clin Cancer Res* 2008; **14**: 3916-3925.

20. Hay T, Matthews JR, Pietzka L*, et al.* Poly(ADP-ribose) polymerase-1 inhibitor treatment regresses autochthonous Brca2/p53-mutant mammary

tumors in vivo and delays tumor relapse in combination with carboplatin. *Cancer Res* 2009; **69**: 3850-3855.

21. Varela I, Klijn C, Stephens PJ*, et al.* Somatic structural rearrangements in genetically engineered mouse mammary tumors. *Genome Biol* 2010; **11**: R100.
22. Wartman LD, Larson DE, Xiang Z*, et al.* Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression. *J Clin Invest* 2011; **121**: 1445-1455.
23. Yuan W, Stawiski E, Janakiraman V*, et al.* Conditional activation of Pik3ca(H1047R) in a knock-in mouse model promotes mammary tumorigenesis and emergence of mutations. *Oncogene* 2013; **32**: 318-326.
24. McFadden DG, Papagiannakopoulos T, Taylor-Weiner A*, et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* 2014; **156**: 1298-1311.
25. Selbert S, Bentley DJ, Melton DW*, et al.* Efficient BLG-Cre mediated gene deletion in the mammary gland. *Transgenic Res* 1998; **7**: 387-396.
26. Jonkers J, Meuwissen R, van der Gulden H*, et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet* 2001; **29**: 418-425.
27. Melchor L, Molyneux G, Mackay A*, et al.* Identification of Cellular and Genetic Drivers of Breast Cancer Heterogeneity in Genetically Engineered Mouse Tumour Models. *J Pathol* 2014.
28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-1760.
29. McKenna A, Hanna M, Banks E*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-1303.

30. Cibulskis K, Lawrence MS, Carter SL*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**: 213-219.

31. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013; **14**: 178-192.

32. Krumm N, Sudmant PH, Ko A*, et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Res* 2012; **22**: 1525-1532.

33. Li J, Lupat R, Amarasinghe KC*, et al.* CONTRA: copy number analysis for targeted resequencing. *Bioinformatics* 2012; **28**: 1307-1313.

34. UCSC In Silico PCR tool. [http://genome.ucsc.edu]

35. MGI curated sets of homology. [http://www.informatics.jax.org/homology.shtml]

36. The Cancer Genome Atlas data portal. [https://tcga-data.nci.nih.gov/tcga/]

37. Farmer H, McCabe N, Lord CJ*, et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005; **434**: 917-921.

38. Bryant HE, Schultz N, Thomas HD*, et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 2005; **434**: 913-917.

39. McNerney ME, Lee KM, Kumar V. 2B4 (CD244) is a non-MHC binding receptor with multiple functions on natural killer cells and CD8+ T cells. *Mol Immunol* 2005; **42**: 489-494.

40. Kandoth C, McLellan MD, Vandin F*, et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 2013; **502**: 333-339.

41. Graveel CR, DeGroot JD, Su Y*, et al.* Met induces diverse mammary carcinomas in mice and is associated with human basal breast cancer. *Proc Natl Acad Sci U S A* 2009; **106**: 12909-12914.

42. Ponzo MG, Lesurf R, Petkiewicz S*, et al.* Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer. *Proc Natl Acad Sci U S A* 2009; **106**: 12903-12908.

43. Knight JF, Lesurf R, Zhao H*, et al.* Met synergizes with p53 loss to induce mammary tumors that possess features of claudin-low breast cancer. *Proc Natl Acad Sci U S A* 2013; **110**: E1301-1310.

44. Smolen GA, Muir B, Mohapatra G*, et al.* Frequent met oncogene amplification in a Brca1/Trp53 mouse model of mammary tumorigenesis. *Cancer Res* 2006; **66**: 3452-3455.

45. Klijn C, Holstege H, de Ridder J*, et al.* Identification of cancer genes using a statistical framework for multiexperiment analysis of nondiscretized array CGH data. *Nucleic Acids Res* 2008; **36**: e13.

46. Liu P, Cheng H, Santiago S*, et al.* Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms. *Nat Med* 2011; **17**: 1116-1120.

47. Xu K, Usary J, Kousis PC*, et al.* Lunatic fringe deficiency cooperates with the Met/Caveolin gene amplicon to induce basal-like breast cancer. *Cancer Cell* 2012; **21**: 626-641.

48. Otto B, Gruner K, Heinlein C*, et al.* Low-grade and high-grade mammary carcinomas in WAP-T transgenic mice are independent entities distinguished by Met expression. *Int J Cancer* 2013; **132**: 1300-1310.

49. Futreal PA, Coin L, Marshall M*, et al.* A census of human cancer genes. *Nat Rev Cancer* 2004; **4**: 177-183.

50. Ciriello G, Sinha R, Hoadley KA*, et al.* The molecular diversity of Luminal A breast tumors. *Breast Cancer Res Treat* 2013; **141**: 409-420.

51. Yoon HG, Chan DW, Reynolds AB*, et al.* N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol Cell* 2003; **12**: 723-734.

52. Fu J, Yoon HG, Qin J*, et al.* Regulation of P-TEFb elongation complex activity by CDK9 acetylation. *Mol Cell Biol* 2007; **27**: 4641-4651.

53. Barbacid M. ras genes. *Annu Rev Biochem* 1987; **56**: 779-827.

54. Stephens PJ, Tarpey PS, Davies H*, et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012; **486**: 400-404.

55. Shah SP, Roth A, Goya R*, et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012; **486**: 395-399.

56. Zhang L, Liu Y, Song F*, et al.* Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. *Proc Natl Acad Sci U S A* 2011; **108**: 13653-13658.

57. Abdul M, Ramlal S, Hoosein N. Ryanodine receptor expression correlates with tumor grade in breast cancer. *Pathol Oncol Res* 2008; **14**: 157-160.

58. Stephens PJ, Davies HR, Mitani Y*, et al.* Whole exome sequencing of adenoid cystic carcinoma. *J Clin Invest* 2013; **123**: 2965-2968.

59. An CH, Kim YR, Kim HS*, et al.* Frameshift mutations of vacuolar protein sorting genes in gastric and colorectal cancers with microsatellite instability. *Hum Pathol* 2012; **43**: 40-47.

60. Thompson BJ, Mathieu J, Sung HH*, et al.* Tumor suppressor properties of the ESCRT-II complex component Vps25 in Drosophila. *Dev Cell* 2005; **9**: 711-720.

61. Vaccari T, Bilder D. The Drosophila tumor suppressor vps25 prevents nonautonomous overproliferation by regulating notch trafficking. *Dev Cell* 2005; **9**: 687-698.

62. Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* 2012; **47**: 497-510.

63. Waters TR, Swann PF. Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *Mutat Res* 2000; **462**: 137-147.

64. Alexandrov LB, Nik-Zainal S, Wedge DC*, et al.* Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 2013; **3**: 246-259.

65. Fong PC, Boss DS, Yap TA*, et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 2009; **361**: 123-134.

66. Sandhu SK, Schelman WR, Wilding G*, et al.* The poly(ADP-ribose) polymerase inhibitor niraparib (MK4827) in BRCA mutation carriers and patients with sporadic cancer: a phase 1 dose-escalation trial. *Lancet Oncol* 2013; **14**: 882-892.

Tables

Gene Symbol	Gene name	Human Ortholog	Sample	Mutation	PROVEAN Prediction
Kras	*v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*	*KRAS*	1001	p.G13R	Deleterious
Notch2	*Notch 2*	*NOTCH2*	919	c.2601+ 6A	N/A
Bcor	*BCL6 interacting co-repressor*	*BCOR*	1035	p.P1099_1110Ydel	Deleterious
Gata3	*GATA binding protein 3*	*GATA3*	1114	p.D6E	Neutral
Tlr4	*Toll-like receptor 4*	*TLR4*	904	p.A132del	Deleterious
Ncor1	*Nuclear receptor co-repressor 1*	*NCOR1*	1114	p.V227A	Neutral

Table 1

Figure and Table legends

Figure 1. *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours are *Brca2* deficient. Tumour and spleen *Brca2* exons 9-12 DNA sequence reads displayed in the Broad Institute Integrative Genomics Viewer (IGV). For each sample the coloured peaks represent the number of DNA sequence reads aligned to the reference genome (mouse mm9) for that genomic position. Black line marks the loss of *Brca2* exon 11 sequence reads in the tumour samples compared to the matched spleens. No loss in sequence read depth is observed in the exons that are not deleted.

Figure 2. Somatic mutations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (A) bar plot of the total somatic mutations in each tumour/spleen comparison. (B) bar plot of the number of base substitutions, deletions and insertions in each tumour/spleen comparison. (C) pie chart of the distribution of somatic mutations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (D) bar plot of the *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour somatic mutations that affect protein coding sequence. (E) bar plot of the somatic mutations that affect protein coding sequence in each tumour/spleen comparison.

Figure 3. Copy number alterations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (A) bar plot of the total CONTRA predicted target region variant gains and losses in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (B) bar plot of the CONTRA predicted target region variant gains and losses in each tumour/spleen comparison. (C) bar plot of the total CoNIFER predicted copy number variant gains and losses in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (D) bar plot of the CoNIFER predicted copy number variant gains and losses in each tumour/spleen comparison. (E)

genomic view of the CONTRA and CoNIFER predicted copy number alterations displayed in IGV. Blue is copy number loss and red is copy number gain.

Figure 4. Somatically mutated genes and copy number alterations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (A) a schematic of the protein coding sequence and known domains of the candidate recurrently mutated genes and genes mutated in human cancer identified in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. The position of mutations are indicated. (B) CONTRA and CoNIFER output for the chromosome 6 region recurrently amplified in the mouse mammary tumours displayed in IGV. Blue is copy number loss and red is copy number gain. (C) confocal microscope images of *Met* FISH on *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour tissue sections. Right panels, a tumour with a predicted *Met* gain has an increase in *Met* probe signal (red). Left panels, a tumour with no predicted *Met* gain. Images are 63x magnification, with inset 100x. Control telomeres are labelled green and nuclei are DAPI stained blue. (D) *Kras* sequence reads from a *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* tumour displayed in the IGV showing tumour DNA contains an alternate base compared to matched normal spleen DNA from the same animal. Right panels show Sanger sequence validation.

Figure 5. Venn diagram comparing *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour mutations to human *BRCA*-associated cancer datasets. a comparison of the somatically mutated genes identified in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours with: (A) human *BRCA2* mutant breast cancer, with 31 genes mutated in both tumour types, (B) human *BRCA2* mutant, *TP53* mutant breast cancer, with 11 genes mutated in both tumour types, (C) human *BRCA2* mutant ovarian cancer, with 16 genes mutated in both tumour types and (D) human *BRCA2*

mutant, *TP53* mutant ovarian cancer, with 10 genes mutated in both tumour types. Genes in blue are recurrently mutated in the human disease.

Figure 6. The spectrum of base substitutions in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (A) frequency of somatic transversion and transition mutations in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mouse mammary tumour exomes. The proportions of base substitutions for each tumour is shown. (B) a line chart illustrating one stable mutational signature was derived from the *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour base substitutions. (C) bar plots of the one reproducible signature of mutational base substitutions extracted from the *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumour DNA exome data. The horizontal axis shows the 96 substitution mutation subtypes. Error bars show standard deviations. The horizontal dotted red line divides the 96 different base substitutions into the most frequent 16 substitutions (above the line) and the remaining 80 substitutions (below the line).

Figure 7. DNA deletions surrounded by microhomology in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (A) total number of exonic sequence deletions flanked by microhomology in each tumour. (B) total number of exonic sequence deletions flanked by microhomology in *Blg-Cre Brca2$^{f/f}$ Trp53$^{f/f}$* mammary tumours. (C) strip chart of the base pair deletion size flanked by microhomology present in the tumours. The red line marks the median. (D) number of deletions and the number of bases flanked by microhomology present in each tumour.

Table 1. Genes with somatic mutations in mouse *Brca2* null, *Trp53* null mammary tumours that are mutated in human cancer. Genes with mutations in protein coding regions and that have human orthologs were compared to genes significantly mutated in human cancer.

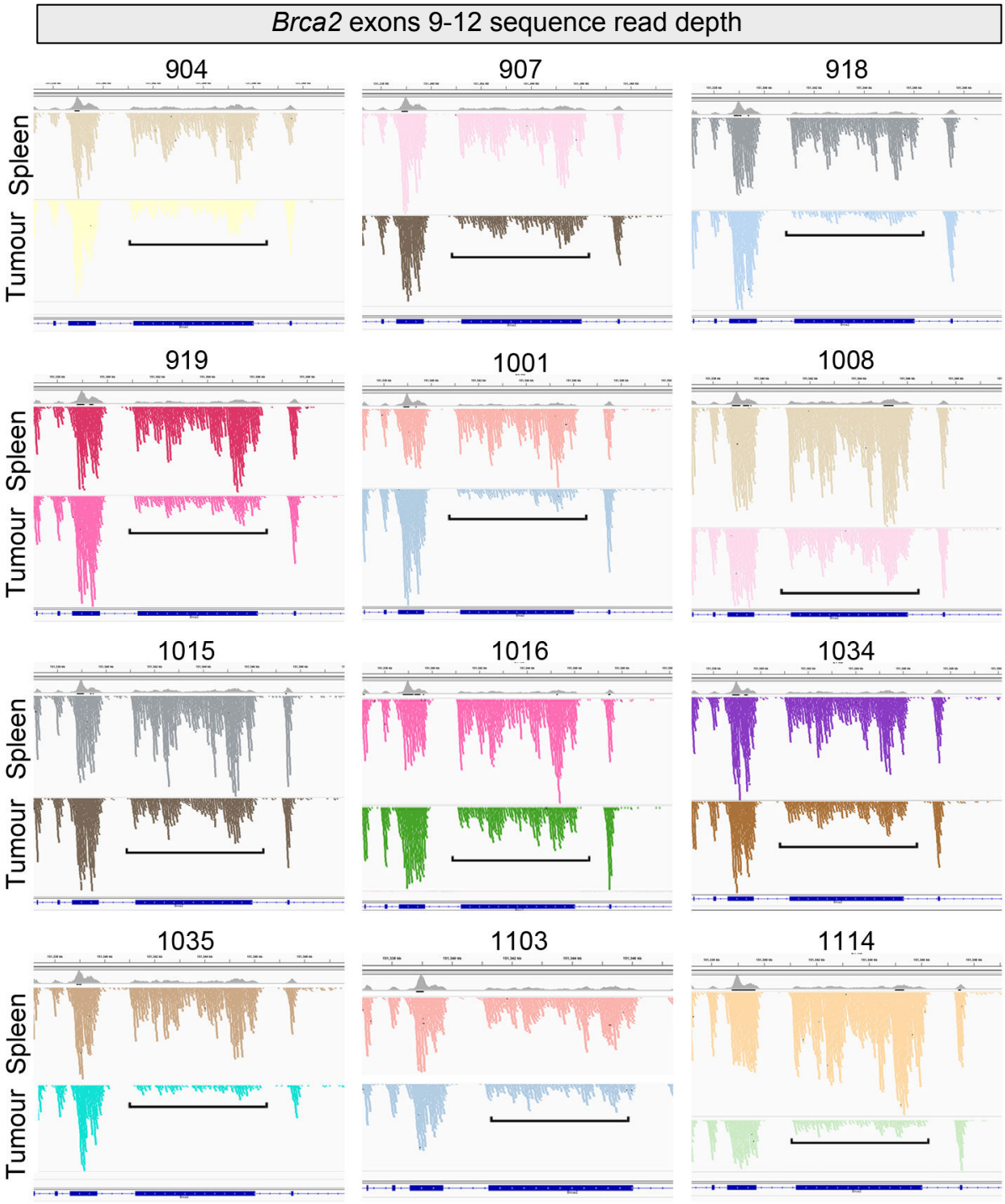

Brca2 exons 9-12 sequence read depth
904
907
918
919
1001
1008
1015
1016
1034
1035
1103
1114
Spleen
Tumour

Figure 1

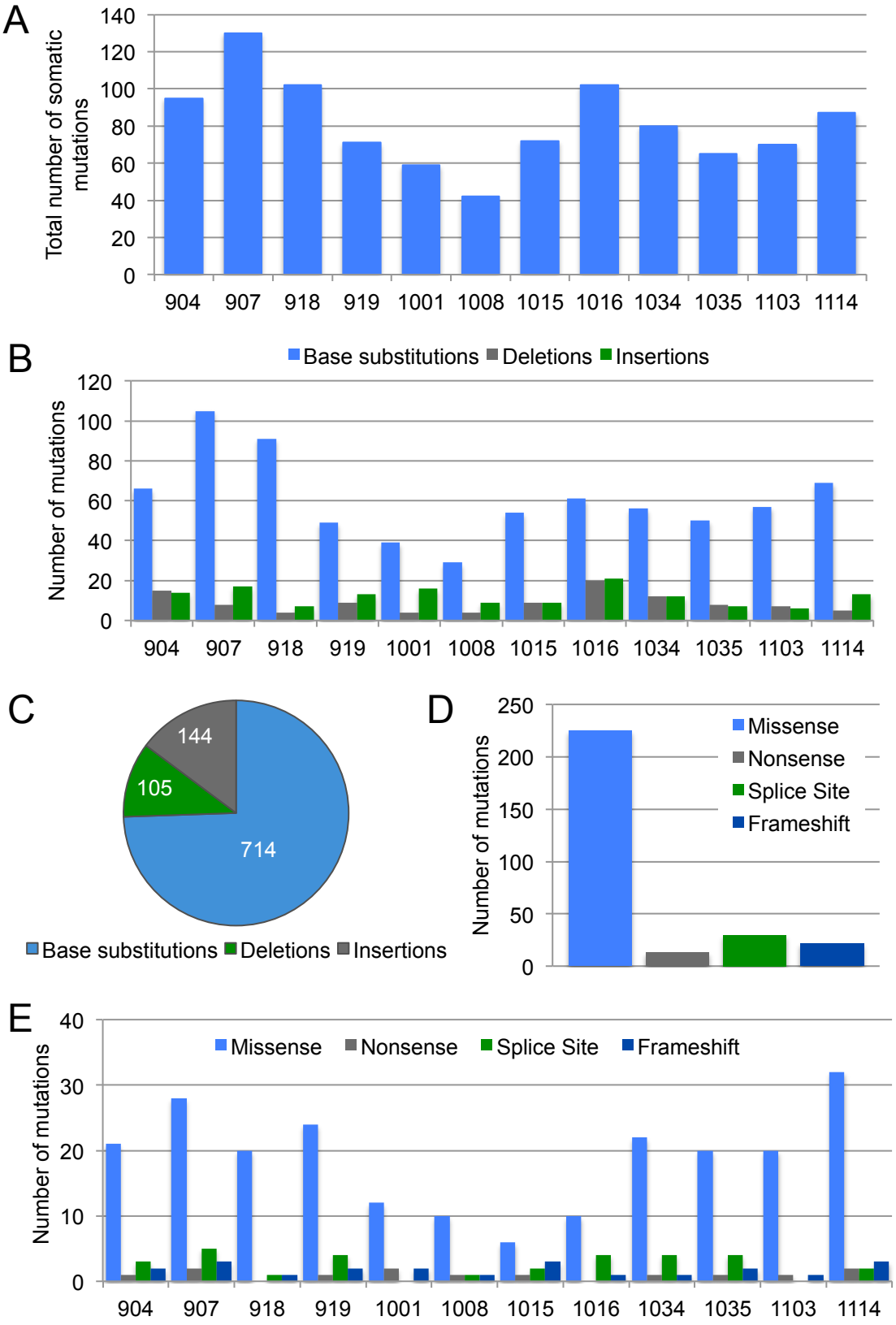

A
Total number of somatic mutations
140
120
100
80
60
40
20
0
904 907 918 919 1001 1008 1015 1016 1034 1035 1103 1114
B
Base substitutions
Deletions
Insertions
Number of mutations
120
100
80
60
40
20
0
904 907 918 919 1001 1008 1015 1016 1034 1035 1103 1114
C
144
105
714
Base substitutions
Deletions
Insertions
D
Number of mutations
250
200
150
100
50
0
Missense
Nonsense
Splice Site
Frameshift
E
Number of mutations
40
30
20
10
0
Missense
Nonsense
Splice Site
Frameshift
904 907 918 919 1001 1008 1015 1016 1034 1035 1103 1114

Figure 2

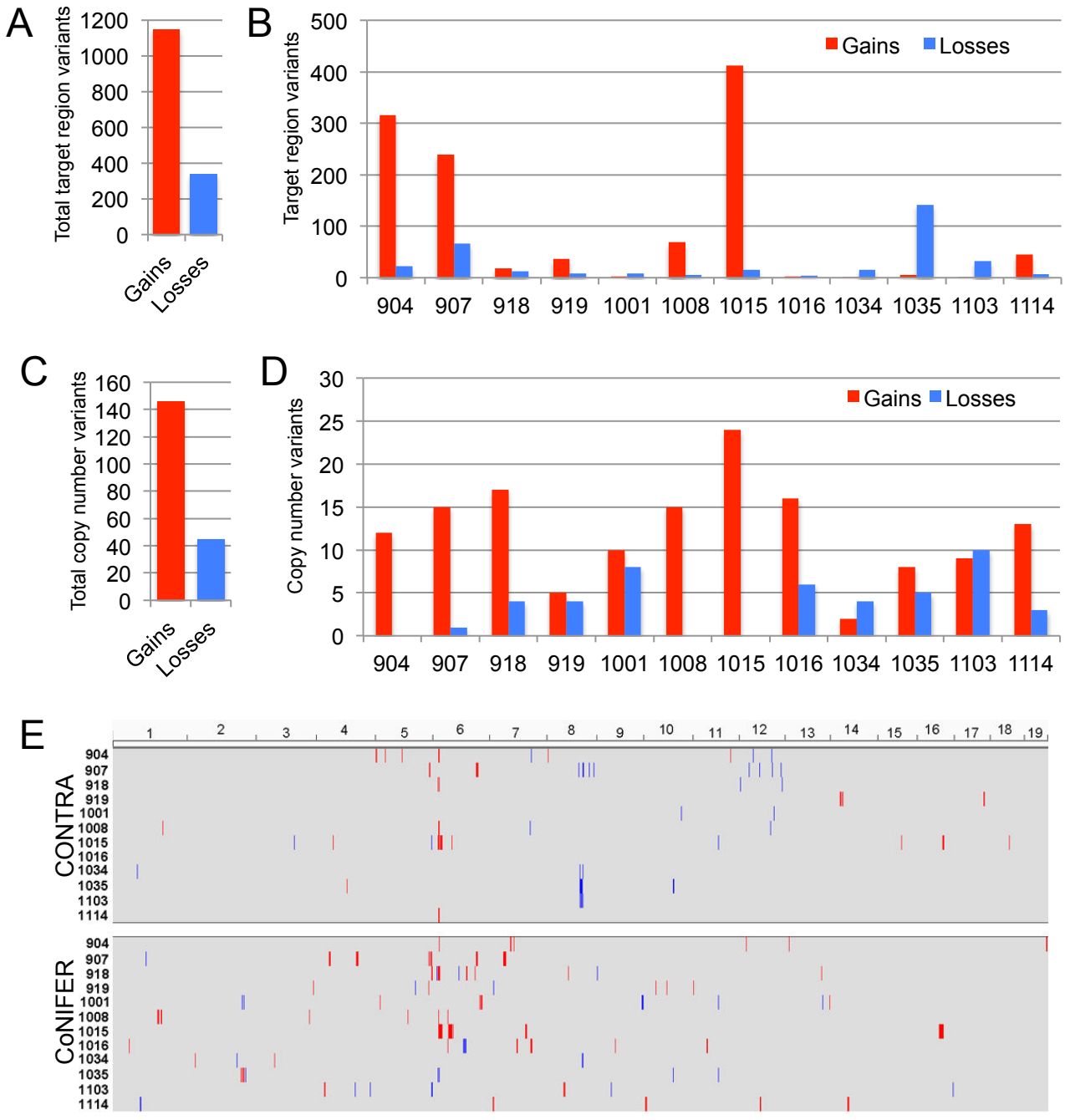
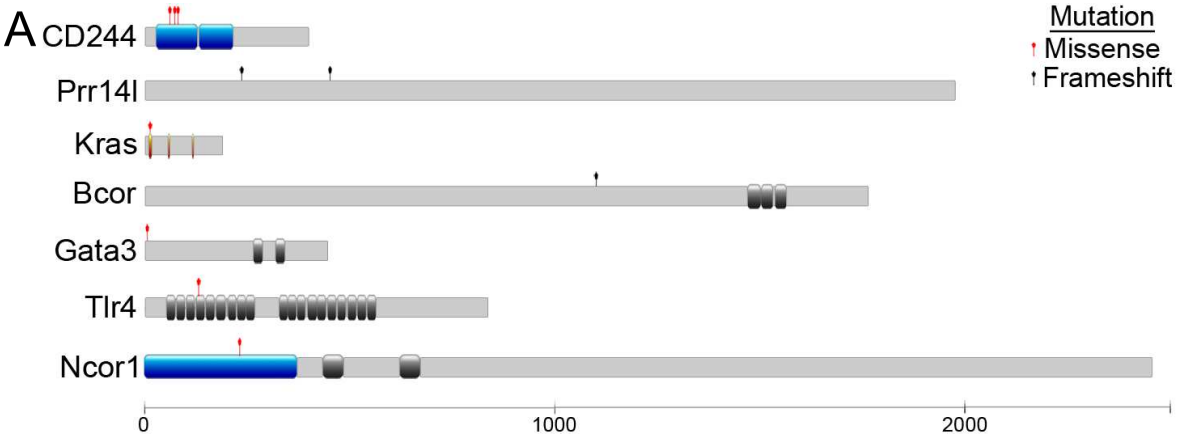
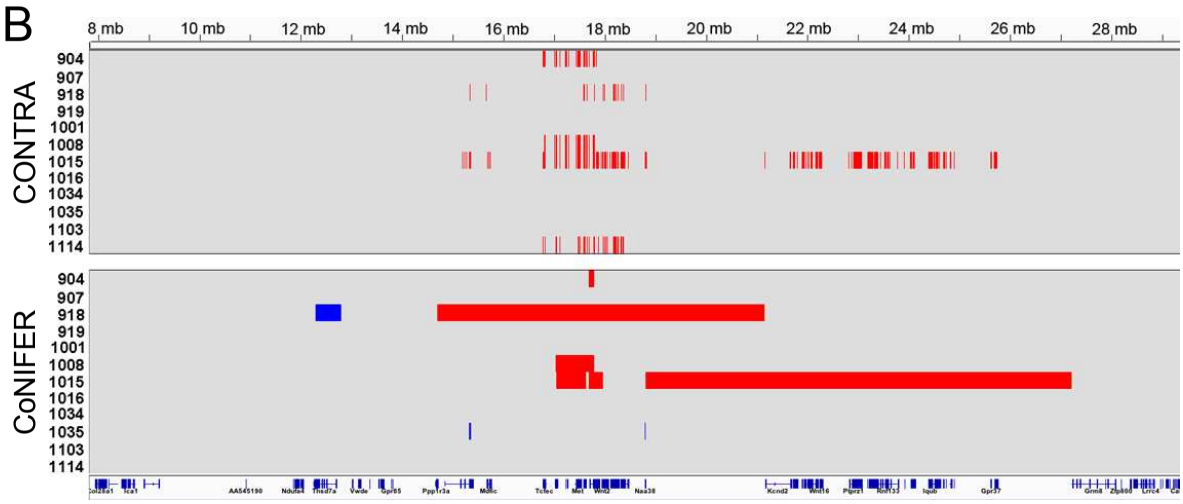
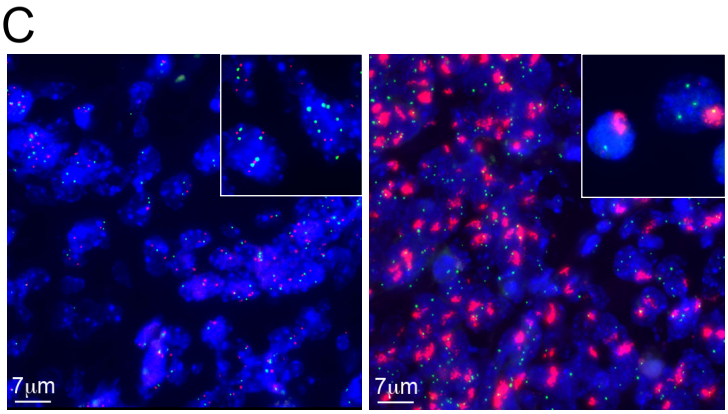
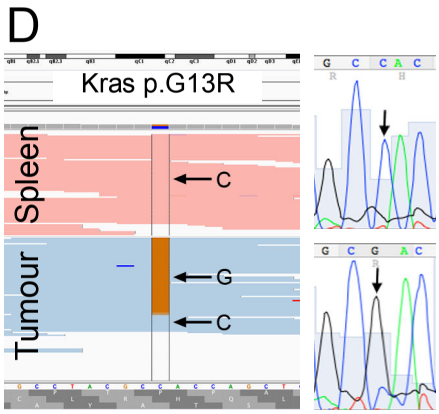

A
Total target region variants
1200
1000
800
600
400
200
0
Gains
Losses
B
Target region variants
500
400
300
200
100
0
Gains
Losses
904
907
918
919
1001
1008
1015
1016
1034
1035
1103
1114
C
Total copy number variants
160
140
120
100
80
60
40
20
0
Gains
Losses
D
Copy number variants
30
25
20
15
10
5
0
Gains
Losses
904
907
918
919
1001
1008
1015
1016
1034
1035
1103
1114
E
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
CONTRA
CoNIFER
904
907
918
919
1001
1008
1015
1016
1034
1035
1103
1114

Figure 3


A
CD244
Prr14l
Kras
Bcor
Gata3
Tlr4
Ncor1
0
1000
2000
Mutation
Missense
Frameshift


B
8 mb
10 mb
12 mb
14 mb
16 mb
18 mb
20 mb
22 mb
24 mb
26 mb
28 mb
CONTRA
CoNIFER
904
907
918
919
1001
1008
1015
1016
1034
1035
1103
1114


C
7μm
7μm


D
Kras p.G13R
Spleen
Tumour
C
G
C

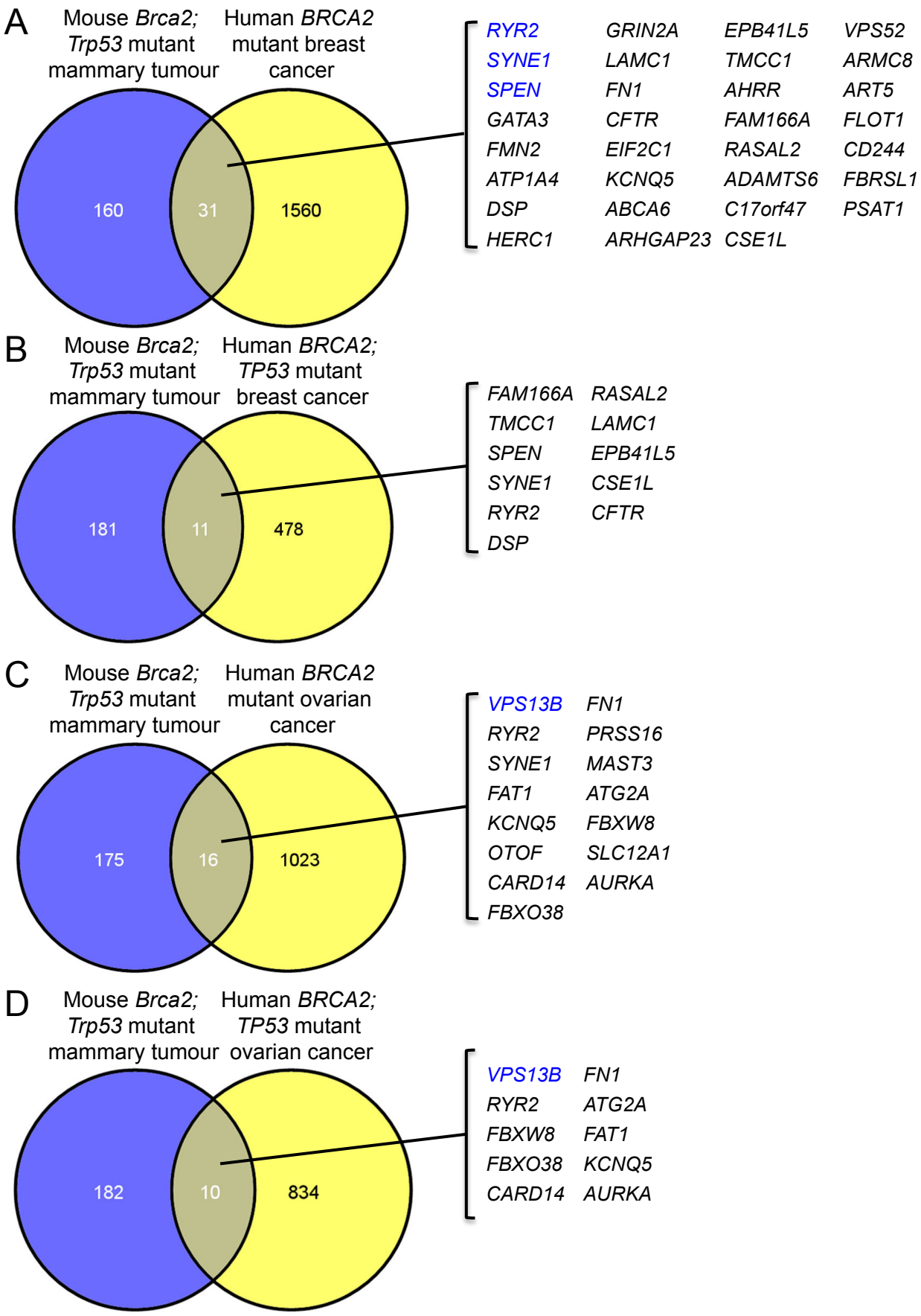
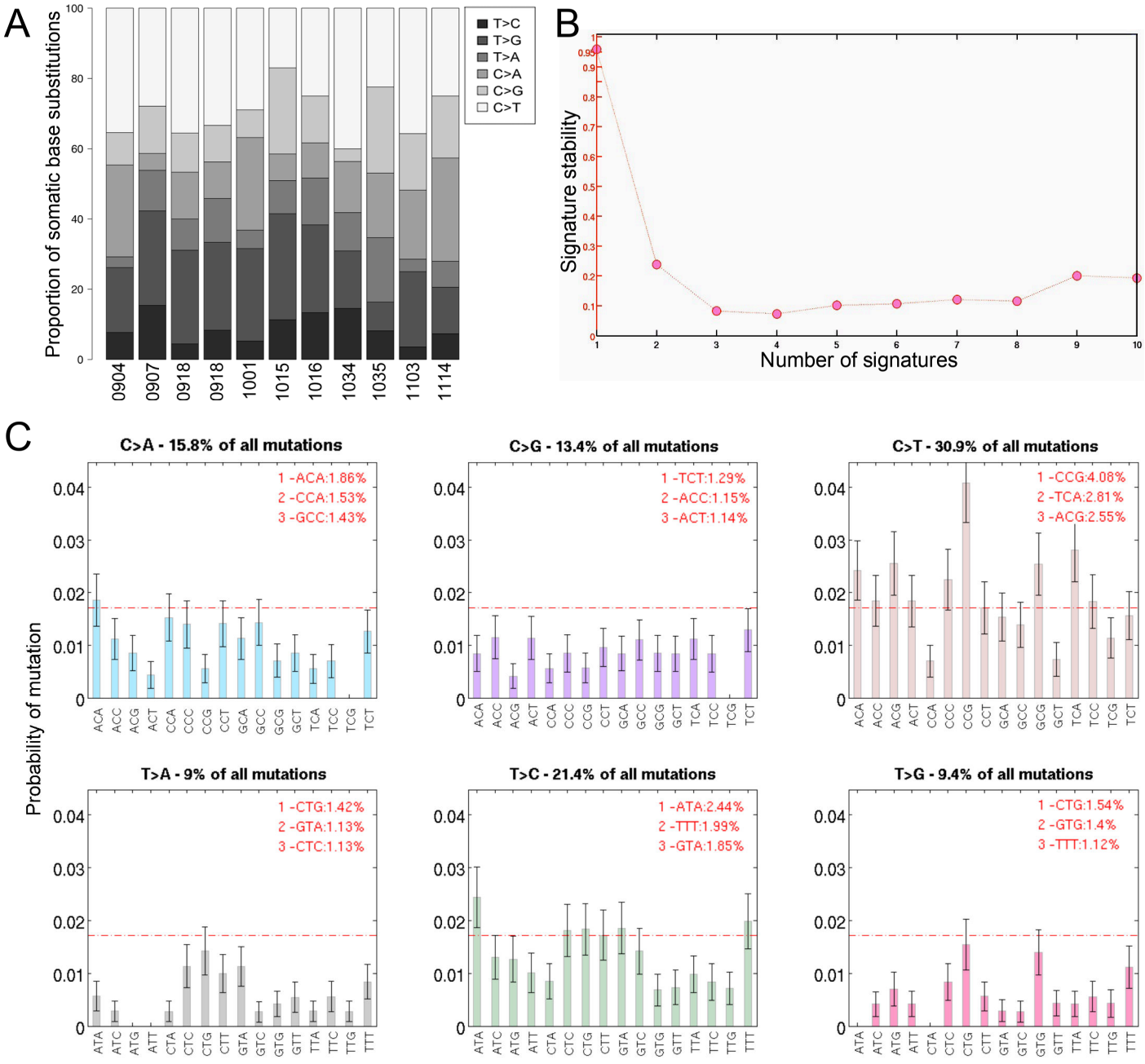
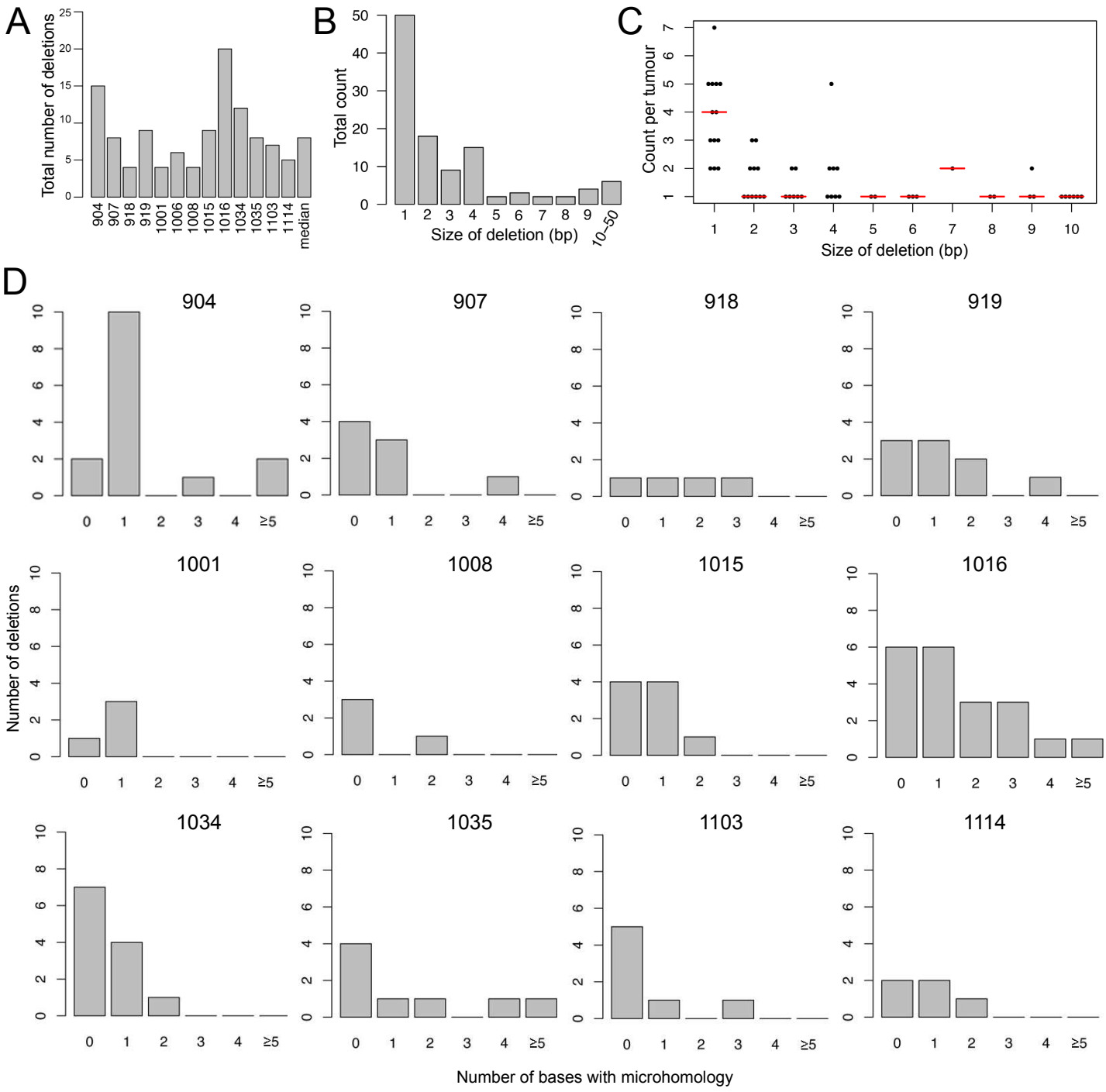

A
Mouse Brca2; Trp53 mutant mammary tumour
Human BRCA2 mutant breast cancer
160
31
1560
RYR2 GRIN2A EPB41L5 VPS52
SYNE1 LAMC1 TMCC1 ARMC8
SPEN FN1 AHRR ART5
GATA3 CFTR FAM166A FLOT1
FMN2 EIF2C1 RASAL2 CD244
ATP1A4 KCNQ5 ADAMTS6 FBRSL1
DSP ABCA6 C17orf47 PSAT1
HERC1 ARHGAP23 CSE1L
B
Mouse Brca2; Trp53 mutant mammary tumour
Human BRCA2; TP53 mutant breast cancer
181
11
478
FAM166A RASAL2
TMCC1 LAMC1
SPEN EPB41L5
SYNE1 CSE1L
RYR2 CFTR
DSP
C
Mouse Brca2; Trp53 mutant mammary tumour
Human BRCA2 mutant ovarian cancer
175
16
1023
VPS13B FN1
RYR2 PRSS16
SYNE1 MAST3
FAT1 ATG2A
KCNQ5 FBXW8
OTOF SLC12A1
CARD14 AURKA
FBXO38
D
Mouse Brca2; Trp53 mutant mammary tumour
Human BRCA2; TP53 mutant ovarian cancer
182
10
834
VPS13B FN1
RYR2 ATG2A
FBXW8 FAT1
FBXO38 KCNQ5
CARD14 AURKA

Figure 5


A
Proportion of somatic base substitutions
T>C
T>G
T>A
C>A
C>G
C>T
0904
0907
0918
0918
1001
1015
1016
1034
1035
1103
1114
B
Signature stability
Number of signatures
C
Probability of mutation
C>A - 15.8% of all mutations
1 -ACA:1.86%
2 -CCA:1.53%
3 -GCC:1.43%
C>G - 13.4% of all mutations
1 -TCT:1.29%
2 -ACC:1.15%
3 -ACT:1.14%
C>T - 30.9% of all mutations
1 -CCG:4.08%
2 -TCA:2.81%
3 -ACG:2.55%
T>A - 9% of all mutations
1 -CTG:1.42%
2 -GTA:1.13%
3 -CTC:1.13%
T>C - 21.4% of all mutations
1 -ATA:2.44%
2 -TTT:1.99%
3 -GTA:1.85%
T>G - 9.4% of all mutations
1 -CTG:1.54%
2 -GTG:1.4%
3 -TTT:1.12%


A
Total number of deletions
0
5
10
15
20
25
904
907
918
919
1001
1006
1008
1015
1016
1034
1035
1103
1114
median
B
Total count
0
10
20
30
40
50
1
2
3
4
5
6
7
8
9
10–50
Size of deletion (bp)
C
Count per tumour
1
2
3
4
5
6
7
1
2
3
4
5
6
7
8
9
10
Size of deletion (bp)
D
904
907
918
919
1001
1008
1015
1016
1034
1035
1103
1114
Number of deletions
0
2
4
6
8
10
0
1
2
3
4
≥5
Number of bases with microhomology