

The Q-IMPROvE (Queensland-Implementation of Precision Oncology in breast cancer) pilot study.

Amy E McCart Reed^{1}, Therese McCurry², Georgina Hollway^{3,4}, Haidar Al-Saig², Vladimir Andelkovic⁵, Katharine Cuff⁵, Margaret Cummings⁷, David Fairbairn⁷, Po-ling Inglis², Gillian Jagger⁵, Helene Johanson⁷, Lauren Kalinowski¹, Olga Kondrashova⁴, Lambros T. Koufariotis⁴, Anna Kuchel^{2,8}, Rahul Ladwa^{5,8}, Chiyun Lau^{7,8}, Ben Lundie⁷, Helen Mar Fan^{8,9}, Nicole McCarthy¹⁰, Kathryn Middleton^{5,11,12}, Kowsalya Murugappan^{2,8}, Mark Nalder^{2,8}, Colleen Niland¹, Michelle K Nottage², Kenneth J O'Byrne^{5,6}, John V Pearson⁴, Kate Roberts^{5,8}, Gorane Santamaria Hormaechea⁵, Cameron Snell^{12,13^}, Karin Steinke², Aneta Suder², Diana Tam², Euan Walpole⁵, Natasha Woodward^{11,12}, Clement Wong², Ho Yi Wong^{3,14}, Wen Xu^{5,8}, Peter T Simpson¹, Nicola Waddell^{3,4}, and Sunil R Lakhani^{1,7*}*

1. The University of Queensland, UQ Centre for Clinical Research, Herston, Qld, Australia
 2. Royal Brisbane and Women's Hospital, Herston, Qld, Australia
 3. genomiQa, Herston, Qld, Australia
 4. QIMR Berghofer, Herston, Qld, Australia
 5. Princess Alexandra Hospital, Woolloongabba, Qld, Australia
 6. Queensland University of Technology, Brisbane, Qld, Australia
 7. Pathology Queensland, Royal Brisbane and Women's Hospital, Herston, Qld, Australia
 8. The University of Queensland, School of Clinical Medicine, Brisbane, Qld, Australia
 9. Genetic Health Queensland, Herston, Qld, Australia
 10. Icon Cancer Care, Wesley Hospital, Auchenflower, Qld, Australia
 11. Mater Hospital Brisbane, South Brisbane, Qld, Australia
 12. Mater Research, University of Queensland, Brisbane, Qld, Australia
 13. Mater Pathology, Brisbane, QLD, Australia; South Brisbane, Qld, Australia
 14. The University of Queensland Diamantina Institute, Woolloongabba, Qld, Australia
- [^] current address Anatomical Pathology, Peter MacCallum Cancer Centre, Melbourne, Vic, Australia.

Corresponding authors:

Sunil R Lakhani: s.lakhani@uq.edu.au

Amy E McCart Reed: amy.reed@uq.edu.au

Keywords: whole genome sequencing, clinical genomics, precision oncology, breast cancer, homologous recombination deficiency, actionable mutations, neoadjuvant setting

Abstract (228 words)

Background: The cancer genomics field has embraced the advent of precision oncology, and vast volumes of data have been mined for biomarkers of drug actionability. While some cancers, such as lung cancer, have detailed panels of actionable genomic biomarkers, sequencing panels have been less useful in breast cancer given its large number of cancer driver genes mutated at a relatively low frequency. Furthermore, mutation signatures have potential to assist in identifying homologous recombination deficient tumours for targeting

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

with PARP inhibitor therapy.

Patients and Methods: To investigate whether whole genome sequencing could benefit breast cancer patients we initiated the Q-IMPROvE (Queensland-IMplementation of PRrecision Oncology in brEast cancer) prospective pilot study. We report the analysis of matched tumour and normal genomes of 28 high-risk breast cancer patients undergoing treatment in the neo-adjuvant setting.

Results: Using whole genome sequencing, we detected actionable events that would otherwise not have been identified. A quarter of patients demonstrated a defect in homologous recombination DNA repair using the HRDetect and HRD scores. Germline variants of importance (*BRCA1*, *CHEK2*) were identified in two patients that did not meet clinical guidelines for germline genetic testing. Somatically, *TP53* and *PIK3CA* were the most commonly mutated breast cancer driver genes.

Conclusions: We have demonstrated the benefit of whole genome sequencing of both the tumour and germline for breast cancer patients otherwise not meeting clinical criteria for genetic health referrals.

Main

Most breast cancer driver genes are mutated in less than 5% of cancers resulting in significant interpatient heterogeneity [1, 2], and thus at the driver gene level, Stephens *et al* consider that most primary breast cancers are distinct [2]. The genomics of breast cancer has been comprehensively studied using whole genome sequencing (WGS) of 560 breast cancers [3]. Saturation analyses predict that based on the number of samples now sequenced, most cancer genes implicated in 2% or more breast cancers will have been identified [4]. Even so, a comprehensive effort by Yates *et al*. [5] acknowledges that many more low-frequency cancer genes remain to be discovered, because there is emerging evidence that metastatic, pre-treated, and special histopathologic subtypes of cancers are genomically distinct from the “general” breast cancer population. Identification of these subtypes allows implementation of specific therapies for the different types of breast cancers, enabling each patient to become an ‘n-of-one’ trial, with their response to therapy informing the treatment of future patients. Increasing numbers of tools are emerging to facilitate the matching of genomic alterations and therapies, including for example, OncoKB [6] and PanDrugs [7], while the MD Anderson program [8] is feeding back ‘sequence-drug’ matching data into the public arena through their Precision Cancer Therapy interface. By end 2015, 39 gene targets with matched FDA-

approved therapies were noted in an extensive review of precision oncology [9]; and, as of 2022, OncoKB notes 43 genes at Level 1 with approval, and 11 resistance genes (Level R1/R2), across all cancer types.

In some cancer types, a small panel of pre-selected genes can be sequenced to identify mutation status, and deduce therapeutic strategies (e.g. lung, colorectal cancer). Whole exome sequencing encompassing all of the genome's coding regions can also be useful, however in cancer types such as breast, it can be less reliable for assessing copy number alterations and will miss structural variations occurring outside of exons [10]. WGS allows us to interrogate gene fusions and larger structural rearrangements [11], and perhaps most importantly for breast cancer, to determine mutation signatures [12]. Certain mutation signatures have potential as diagnostic tools. For example, mutation signatures associated with a defective homologous recombination (HR) DNA repair pathway, which is crucial for maintaining genome stability, can be more robust at predicting HR deficiency (HRD) than the analysis of genetic alterations in HR genes alone (e.g. *BRCA1*, *BRCA2*, *RAD51C*, *RAD51D*, *PALB2* etc) [13]. HRD enhances sensitivity to DNA-damaging chemotherapeutics, making the associated mutation signature a predictive biomarker. The HRDetect tool [14] was developed in breast cancer and combines several types of mutational signatures derived from WGS to predict HR deficiency (HRD) and thus sensitivity to DNA-damaging chemotherapy (for example anthracyclines or platinum-based therapies). There is increasing evidence of HRDetect score applicability in the clinic; it is tolerant of both low tumour cellularity (as low as 13%) and high mutation burden, and proven in the neoadjuvant setting [14, 15]. Moreover, the OlympiA trial has confirmed that in high-risk early breast cancer (HER2-negative breast cancer and germline *BRCA1* or *BRCA2* pathogenic or likely pathogenic variants) there is a significant improvement in disease free survival with the adjuvant addition of the PARP inhibitor, olaparib [16]. Indeed, the importance of germline variants in seemingly non-familial cancer patients is just coming to light, and expanded germline DNA testing in the Memorial Sloan Kettering breast cancer cohort not meeting clinical criteria for germline testing identified a pathogenic or likely pathogenic germline variant in 17.5% cases [17].

We present a pilot study to examine the potential benefit of the implementation of WGS in the breast cancer care pathway in Queensland, Australia.

Methods

Ethical approval for this pilot study was granted by the Royal Brisbane and Women's Hospital (HREC/2019/QRBW/48171) and The University of Queensland (2020000203). The registered study (ACTRN12621001285842) was funded as part of Queensland Genomics, a state-wide program to implement genomics into public healthcare [18]. Following a diagnosis of breast cancer, and a decision to undergo neo-adjuvant chemotherapy (as per local standard of care protocols), patients provided informed consent and were recruited to the study. At the time of surgical clip insertion, three 16-gauge core biopsies were taken and collected into RNALater (ThermoFisher Scientific, Melbourne, Australia). DNA and RNA were extracted from tumour cores using the AllPrep kit (Qiagen, Melbourne, Australia). DNA from baseline bloods collected in EDTA tubes was extracted using either the chemagic360™ instrument (PerkinElmer, ThermoFisher Scientific) and Janus extraction kit (PerkinElmer) or the QIASymphony instrument and dedicated kit (Qiagen).

Samples were sequenced by BGI Australia using PE150 chemistry on a DNBSeg-G400 sequencer to a targeted minimum read depth of 30x coverage for normal and 60x for tumour DNA. Comprehensive variant analysis was performed by genomIQa (Brisbane, Australia). Sequence data was aligned to the human genome (GRCh37/hg19) using BWA-MEM, single nucleotide variants (SNVs) and indels were identified using GATK [19] and qSNP [20]. Copy number alterations (CNAs) and tumour content were called using ASCAT [21]. Variants were classified using the ACMG [22] and AMP/ASCO/CAP [23] guidelines. Reporting of somatic variants was restricted to n=261 genes selected based on a literature review including reported breast and other cancer driver genes [3, 24, 25], and germline variants to 6 genes based on current local standard for testing and thus clinically actionable (**SupTable 1**). Pharmacogenomic loci assessed in the germline sequences for each patient were also noted [26].

SNV mutational signature assignment was performed using deconstructSigs [27] with version2 COSMIC signatures and a signature contribution cut-off of 10%. Rearrangement signature (RS) assignment was performed using YAPSA [28] with previously reported RS signatures [3] and signature contribution cut-off of 10%. Homologous recombination deficiency (HRD) was estimated using scarHRD [29] using allele-specific copy number information determined by ascatNGS, and by HRDetect [14] using HRD sum scores, insertions

and deletions with microhomology, SNV and RS signatures. An HRD sum score of ≥ 42 and HRdetect score of >0.7 were used to categorise cancers as HRD. A gene is considered amplified if there are 6 copies or more, or in cases where the overall ploidy is 3.5x or greater, a gene must be amplified 2.5x above the ploidy.

Genomic data was presented as a variant report at regular Molecular Tumour Board (MTB) meetings. We aimed to have representatives from each of the following disciplines present; pathology, molecular oncology, clinical genetics, surgery, imaging, data analysts, cancer genomics specialists, with the requirement that the patient's medical oncologist (or their representative) be present.

Results

We present a pilot study and infrastructure pipeline for the clinical implementation of whole genome sequencing for breast cancer patients at three hospitals in Brisbane, Australia.

High-risk breast cancer patients due to undergo neo-adjuvant chemotherapy for their breast cancer were recruited to the study and had a blood sample and research core biopsy taken at the time of surgical clip insertion (prior to the commencement of their chemotherapy). These samples underwent WGS, data were analysed and a comprehensive variant report prepared (germline and somatic variants, mutation signatures, HRD scores, pharmacogenomics) and discussed at a multidisciplinary molecular tumour board (MTB) meeting (**Fig1**). Of 29 patients recruited, 28 patients produced sufficient tumour and germline DNA for sequencing, and 26 tumour-germline genome pairs passed quality assurance and proceeded to a variant report. The variant report discussion centred around the identification of potential germline variants that impacted tumour predisposition and known adverse effects to certain treatments (pharmacogenomics), as well as somatic variants that would be considered targetable in the case that the patient required a second line therapy (lack of response to therapy, or a disease recurrence in future). This was discussed in the context of breast cancer biomarkers Estrogen Receptor (ER) and Human Epidermal Growth Factor Receptor (HER2), as indicated from the core biopsy pathology information, and were classed as ER positive (ERpos) or negative (ERneg) or with or without amplification of HER2 (HER2pos, HER2 neg, respectively).

There were 10 ERpos/HER2neg, five ERpos/HER2pos, seven ERneg/HER2pos and seven TNBC (triple negative breast cancer; ERneg/Progesterone Receptor neg/HER2neg) and the patients

ranged in age from 26-69 (median 49), thus reflecting the local clinical guidelines for neo-adjuvant therapy [30], (**Fig2A**). As shown in **Fig2B**, 14/29 cases had a pathologic complete response (pCR) and 15/29 had residual cancer. All ERpos/HER2neg cases had residual cancer (**Fig2C**) and there was a significant enrichment for HER2pos cases in those with pCR (P=0.0008; chi square).

Of the 29 patients recruited to the study in the funding-mandated timeline, 28 samples were successfully processed into tumour DNA/RNA and germline DNA. The sequencing metrics are detailed in **SupTable 2**. Briefly, the tumour samples were sequenced to an average depth of 73 (55.8 to 90.3); and the normal to 36.5 (33.6 to 40.6). Following sequencing and analysis, 26 samples proceeded with analysis and were prepared into a variant report for discussion, with two samples excluded due to low tumour content.

The importance of germline sequencing

An important feature of our study was the germline sequencing, which is not always performed in diagnostic tumour genomic testing. To adhere to our strict ethical oversight, we filtered germline alterations to report only those considered clinically actionable (**SupTable 1**). Eight patients had an existing referral for clinical genetics testing and counselling (Genetic Health Queensland; local clinical guidelines for referral applied and included the Manchester score [31]), two of these patients were known to be *BRCA1* mutation carriers, the other six patients had no known pathogenic alteration (**SupTable 3**). Germline WGS detected the two known *BRCA1* mutations, in addition, two patients were referred to Genetic Health Queensland after Q-IMPROvE WGS analysis identified pathogenic germline *BRCA1* and a *CHEK2* mutation (with copy neutral loss of heterozygosity in the tumour), demonstrating the value of germline sequencing with restricted analysis. An additional patient was also referred to Genetic Health Queensland for accredited germline genetic testing due to a young age of diagnosis and the detection of a somatic *TP53* (C135Y) variant and a copy neutral LOH; the presence of a germline variation would impact the decision to go forward with radiation therapy in the short term, as well as family cancer implications.

Pharmacogenomics is an emerging field that can facilitate the appropriate targeting of existing drugs to patients, in a manner which will reduce potential toxic side effects identified by specific germline variations. Most commonly, a patient's predisposition to toxicity from capecitabine/fluorouracil can be predicted by investigating the genotype of the *DPYD* gene.

Typically, 10-40% of patients show a severe toxic response to this class of chemotherapeutic [32], and we identified one patient with a heterozygous variation known to impact drug metabolism and mediate toxicity (rs55886062; c.1679 T>G). Crucially this patient had an incomplete response to therapy with residual cancer remaining, and the sequence data informed the selection of an alternative chemotherapeutic to capecitabine, in order to reduce potential toxicity during second-line therapy. We also assessed a number of pharmacogenomic loci as detailed in Low *et al* [26]. However, this data did not yield informative outcomes, for example, three clinically HER2 negative patients had genotypes consistent with trastuzumab/lapatinib sensitivity. While 16 patients (57%) were predicted to have variants associated with neuropathies, our clinical team noted anecdotally that neuropathy is a very common chemotherapy induced side effect, and that these variants would be unlikely to change their management of the patient at this stage.

In summary, the germline analysis identified an additional three instances of actionable alterations, and independently confirmed two existing germline mutations.

Somatic Alterations: what did we look for, what did we find?

A subtractive analysis of tumour compared to normal was performed to characterise the somatic alterations in each patient (**Fig3**). In addition to mutations, the tumour genomes were assessed for mutation burden, ploidy, mutational signatures, chromosomal rearrangements and copy number alterations. As expected, these breast cancer biopsies did not show a high tumour mutation burden, with a range of 0.4 – 5.7 mutations/Mb (median 1.6), thus none met the 10 mutations/Mb required for predicting pembrolizumab eligibility. The range of ploidy status was 1.6-5, with a median of 3.16.

We curated a list of 261 breast cancer relevant genes to inform the somatic nucleotide variants and indels included on the report (**SupTable 1**). *TP53* was the most frequently mutated gene (n=13), followed by *PIK3CA* (n=12 across 10 patients), *MAP3K1* (n=4) and *BRCA1* (n=3) (**Fig4A**). *TP53* variations were spread throughout the gene (**Fig4B**) and included three intronic/splice site mutations; these variants were considered as likely oncogenic but of unknown significance in terms of actionability [33]. *PIK3CA* mutations were predictably clustered, with seven at residue 1047 (**Fig4C**); these were classed as a Tier IA mutation classification in ERpos patients.

With respect to copy number alterations, **Fig4D** shows the large spread of tumour suppressor genes (n=36) with a single copy loss in two or more patients. An additional 35 genes (**SupTable 4**) demonstrated single copy loss in individual patients only. Copy neutral loss of heterozygosity, as a second hit, was most frequently found in *BRCA1* (n=5 patients) and *TP53* (n=3) as shown in **Fig4E (SupTable 4)**. *CDKN2A* (p16^{INK4a}) and *CDKN2B* (p15^{INK4b}), which co-localise on 9p21.3, were found to be homozygously deleted in a single patient, and it is likely that the loss of these well-characterised tumour suppressors impacted tumour development. Frequently amplified genes included *MYC* (n=5 patients), *GATA3* (n=5), the actionable biomarker *CCND1* (n=3), and *CCNE1* (n=2), *MYB* (n=2) and *PIK3CA* (n=2) (**Fig4F; Table 2**) with a high gain reported in an additional 11 genes in individual cases only. *ERBB2* (HER2) was the most recurrently amplified (n=10), the *ERBB2* amplifications were consistent with those cases clinically reported as HER2 amplified by *in situ* hybridisation (**Fig4G**), with the highest *ERBB2* copy number reported to be 83 copies.

Homologous Recombination Deficiency analysis

Mutation signature analysis was used to estimate HRD using HRD Score [34] and HRDetect [14], which are two clinically implementable algorithms used to measure HRD as a proxy for *BRCA1*, *BRCA2* or other key HR gene dysfunction. These algorithms are incredibly useful, as there are many modes of inactivation of *BRCA1/2*, including mutation, large-scale rearrangements, copy number alterations and promoter methylation; not all variations are known to be pathogenic and no single test can capture all of these potential (epi-)genotypes. For Q-IMPROvE, we required that the criteria for both HRD Score and HRDetect be met to class a patient as HRD. In total, seven cases were determined to be HRD (**Fig5**). Four of these cases had a germline line alteration in *BRCA1/CHEK2*, there was a single somatic *BRCA1* mutation (second/first hit unknown), and two cases of unknown aetiology. In the cohort, there were four cases meeting the HRD Score cut-off (>42) but not the HR Detect (>0.7) and so were classed as HR proficient, but no cases with the converse (**Fig5B**).

Somatic sequencing informed actionability

Taken together, we found 18 actionable events across 18 patients (69%) through tumour sequencing of 26 patients. Considering currently actionable changes, **Fig5C** summarises the association between those genomic biomarkers and the patient outcome overall; indicating

that actionable events may prove important in the instances of tumour recurrence. Notably, an incomplete response to neo-adjuvant therapy in tumours where *PIK3CA* were identified (**Fig5D**) indicating that these patients could respond to alpelisib as a second line therapy to target the *PIK3CA* mutations should they relapse. Another actionable change identified was *CCND1* amplification (n=3), where patients could be triaged to Palbociclib and Avelumab, through a local clinical trial the MoST 10 substudies 23-24 (ACTRN12620000568910). Four patients with residual cancer showed defective homologous recombination DNA repair, indicating that they may be sensitive to olaparib.

Discussion

We present a clinical framework for the implementation of whole genome sequencing in breast cancer care. We selected tumour-normal paired WGS and accepted its cost limitations in order to prioritise data generation, in particular around germline variants predicting risk and pharmacogenomic toxicities, broad driver gene detection and mutation signature assessment, in particular for HRD assessment. Furthermore, we necessitated the inclusion of paired normal sequencing to eliminate uncertainty around the identification potential germline findings in somatic data. The traditionally long turnaround time of WGS and analysis was offset by our application of the technology in the neo-adjuvant setting, where sequencing and analysis could occur during treatment cycles, which typically take 20-24 weeks [30].

We identified seven cases of HRD in total, with only two of these cases commencing the study with a confirmed pathogenic alteration in a germline breast cancer predisposition gene. Regarding gene specific actionability, we present 12 *PIK3CA* mutations harboured by ten patients; a single case of *DPYD* predicting toxicity; and, three cases of *CCND1* amplification. We confirmed that WGS-derived *ERBB2* copy number status could reliably replace HER2 IHC and *in situ* hybridisation. In 9/28 patients, no actionable changes were identified. While no therapy changes were mandated by our study, it is clear that there is the capacity for WGS to be applied with outcomes benefiting patients around therapy selection, and clinical trial engagement.

That seven of 28 patients (25%) showed HRD, is a strong argument for its routine clinical implementation in breast cancer care. Our application of a virtual panel filter of the germline sequence to restrict findings to those of breast cancer relevance was ethical, useful, and readily implementable. Furthermore, as risk genes, pharmacogenomics and the mutation signature space are all dynamic with novel data likely to be emerging, a whole genome sequence can be reanalysed to include any new loci of interest, representing a future proofing of the data.

We acknowledge the limitations of this study, in its small size and wholly neo-adjuvant setting. Additionally, we accept that the genome of an early breast cancer, pre-treatment, is unlikely to be identical to the genome of a post-treatment tumour being discussed for second-line therapy. However, we considered that the decision to trial WGS in neoadjuvant patients outweighed the potential problems presented by therapy-induced genome alterations.

This study provides evidence to support the introduction of genomics as standard of care in breast cancer management in Australia. Further study is required to determine whether a custom panel or similar could recapitulate this level of data generation and improved cost and turnaround time, and whether the adjuvant treatment setting would also be appropriate.

Acknowledgements

We acknowledge funding from Queensland Genomics (Queensland Health) and the Medical Research Futures Fund, Genomics Health Futures Mission. Thank you to the patients and their families. We thank the many staff across the RBWH, PAH and Mater hospitals that helped facilitate this study. The study was funded by Queensland Health through Queensland Genomics as an Innovation study.

Author contributions.

Study design and infrastructure: AMR, CS, EW, KOB, NM, NWA, NWO, PTS, SRL.

Patient recruitment co-ordination: GJ, TM

Patient recruitment, clinical oversight, and MTB input: AK, AS, CS, CW, DT, EW, HA, HM, KC, KMi, KMu, KOB, KR, MNa, MNo, NM, NWO, PI, RL, SRL, TM, VA, WX.

Biopsy collection: GSH, KS.

Biopsy review: MC, SRL.

Sample processing: HJ.

Pathology oversight and interpretation: BL, CL, CS, DF, LK, MC, SRL.

Variant report design: AMR, GH, HM, LK, NWA.

Data analysis: BL, GH, OK, LK, HYW, JVP, PTS, NWA, AMR.

Critical appraisal, manuscript writing, correction, and approval of final manuscript: all authors.

Competing interests.

OK has consulted for XING Technologies. JVP and NWA are founders and shareholders of genomiQa Pty Ltd, and members of its board. GH is the clinical genomics lead at genomiQa Pty Ltd.

References

1. Ellis, M.J., et al., *Whole-genome analysis informs breast cancer response to aromatase inhibition*. *Nature*, 2012. **486**(7403): p. 353-60.
2. Stephens, P.J., et al., *The landscape of cancer genes and mutational processes in breast cancer*. *Nature*, 2012. **486**(7403): p. 400-4.
3. Nik-Zainal, S., et al., *Landscape of somatic mutations in 560 breast cancer whole-genome sequences*. *Nature*, 2016. **534**(7605): p. 47-54.
4. Lawrence, M.S., et al., *Discovery and saturation analysis of cancer genes across 21 tumour types*. *Nature*, 2014. **505**(7484): p. 495-501.
5. Yates, L.R. and C. Desmedt, *Translational Genomics: Practical Applications of the Genomic Revolution in Breast Cancer*. *Clin Cancer Res*, 2017. **23**(11): p. 2630-2639.
6. Chakravarty, D., et al., *OncoKB: A Precision Oncology Knowledge Base*. *JCO Precis Oncol*, 2017. **2017**.
7. Pineiro-Yanez, E., et al., *PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data*. *Genome Med*, 2018. **10**(1): p. 41.
8. Dumbrava, E.I. and F. Meric-Bernstam, *Personalized cancer therapy-leveraging a knowledge base for clinical decision-making*. *Cold Spring Harb Mol Case Stud*, 2018. **4**(2).
9. Meric-Bernstam, F., et al., *A decision support framework for genomically informed investigational cancer therapy*. *J Natl Cancer Inst*, 2015. **107**(7).
10. Ozden, F., C. Alkan, and A.E. Cicek, *Polishing copy number variant calls on exome sequencing data via deep learning*. *Genome Res*, 2022. **32**(6): p. 1170-1182.
11. Li, Y., et al., *Patterns of somatic structural variation in human cancer genomes*. *Nature*, 2020. **578**(7793): p. 112-121.
12. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer*. *Nature*, 2013. **500**(7463): p. 415-21.
13. Nones, K., et al., *Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers*. *Ann Oncol*, 2019. **30**(7): p. 1071-1079.
14. Davies, H., et al., *HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures*. *Nat Med*, 2017. **23**(4): p. 517-525.
15. Zhao, E.Y., et al., *Homologous Recombination Deficiency and Platinum-Based Therapy Outcomes in Advanced Breast Cancer*. *Clin Cancer Res*, 2017. **23**(24): p. 7521-7530.
16. Tutt, A.N.J., et al., *Adjuvant Olaparib for Patients with BRCA1- or BRCA2-Mutated Breast Cancer*. *N Engl J Med*, 2021. **384**(25): p. 2394-2405.
17. Ceyhan-Birsoy, O., et al., *Diagnostic yield and clinical relevance of expanded genetic testing for cancer patients*. *Genome Med*, 2022. **14**(1): p. 92.

18. Vidgen, M.E., et al., *Queensland Genomics: an adaptive approach for integrating genomics into a public healthcare system*. NPJ Genom Med, 2021. **6**(1): p. 71.
19. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. Genome Res, 2010. **20**(9): p. 1297-303.
20. Kassahn, K.S., et al., *Somatic point mutation calling in low cellularity tumors*. PLoS One, 2013. **8**(11): p. e74380.
21. Raine, K.M., et al., *ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data*. Curr Protoc Bioinformatics, 2016. **56**: p. 15 9 1-15 9 17.
22. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology*. Genet Med, 2015. **17**(5): p. 405-24.
23. Li, M.M., et al., *Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists*. J Mol Diagn, 2017. **19**(1): p. 4-23.
24. Michailidou, K., et al., *Association analysis identifies 65 new breast cancer risk loci*. Nature, 2017. **551**(7678): p. 92-94.
25. Bailey, M.H., et al., *Comprehensive Characterization of Cancer Driver Genes and Mutations*. Cell, 2018. **173**(2): p. 371-385 e18.
26. Low, S.K., H. Zembutsu, and Y. Nakamura, *Breast cancer: The translation of big genomic data to cancer precision medicine*. Cancer Sci, 2018. **109**(3): p. 497-506.
27. Rosenthal, R., et al., *DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution*. Genome Biol, 2016. **17**: p. 31.
28. Hubschmann, D., et al., *Analysis of mutational signatures with yet another package for signature analysis*. Genes Chromosomes Cancer, 2021. **60**(5): p. 314-331.
29. Sztupinszki, Z., et al., *Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer*. NPJ Breast Cancer, 2018. **4**: p. 16.
30. eviQ. <https://www.eviq.org.au/medical-oncology/breast/neoadjuvant>
31. Evans, D.G., et al., *Pathology update to the Manchester Scoring System based on testing in over 4000 families*. J Med Genet, 2017. **54**(10): p. 674-681.
32. Amstutz, U., et al., *Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update*. Clin Pharmacol Ther, 2018. **103**(2): p. 210-216.
33. Horak, P., et al., *Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC)*. Genet Med, 2022. **24**(5): p. 986-998.
34. Telli, M.L., et al., *Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer*. Clin Cancer Res, 2016. **22**(15): p. 3764-73.

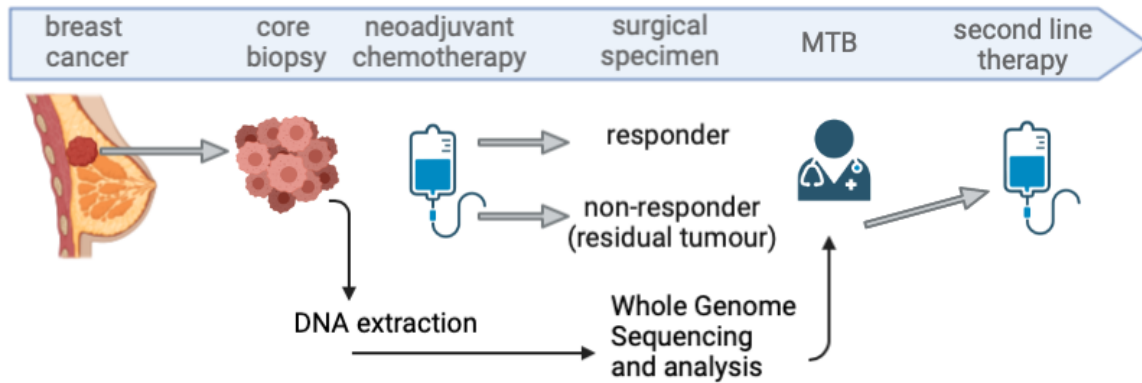


Figure 1: The Q-IMPROvE framework. Following a diagnosis of breast cancer and a decision to proceed with neo-adjuvant therapy, an additional sample (core biopsy) is taken at the time of surgical clip insertion. While the patient undergoes chemotherapy, the pre-therapy tissue and blood sample are sent for DNA extraction and whole genome sequencing. The data is then analysed and returned for discussion at a Molecular Tumour Board (MTB) meeting. Created with BioRender.com.

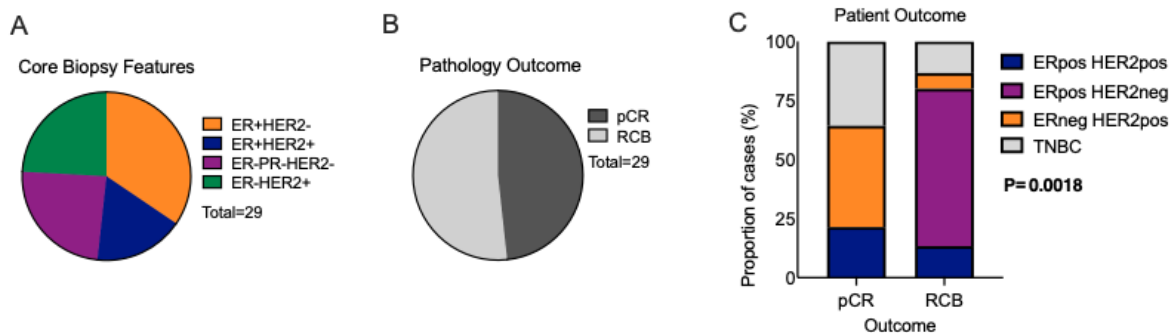


Figure 2: Breast cancer subtypes and outcomes. (A) Of the 29 Q-IMPROvE patients, 11 were HER2pos, there were 15 ERpos and 7 TNBC. (B) 15/29 patients did not achieve a pathologic complete response and had residual cancer burden. (C) The ER pos /HER2neg patients did not respond well to neo-adjuvant therapy and were significantly associated with residual cancer ($P=0.0008$, chi square). pCR, pathologic complete response; TNBC, triple negative breast cancer.

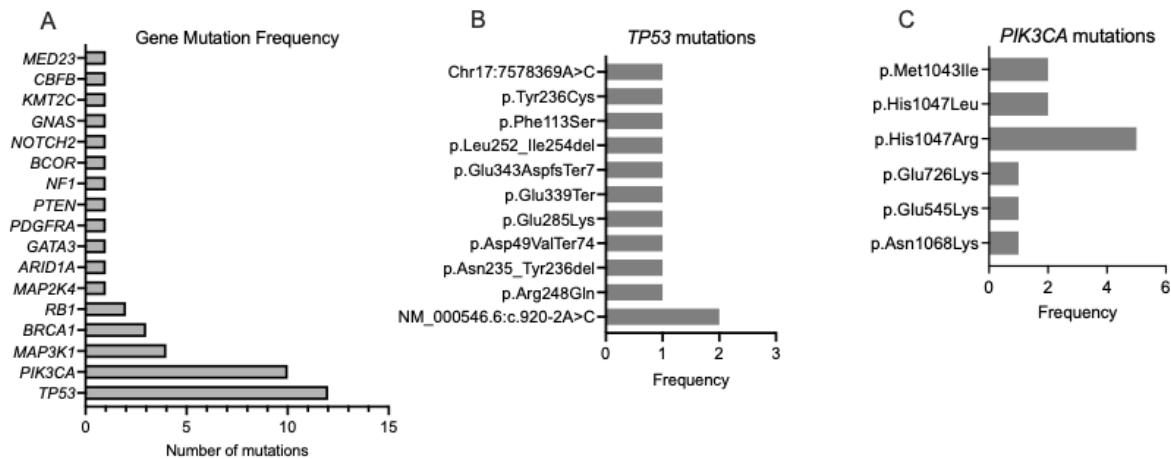


Figure 3: Mutational landscape of Q-IMPROVe cohort. Data per patient, ordered according to standard ER, PR and HER2 status. TMB, tumour mutation burden (mutations/MB). Purity, % tumour cellularity. Ploidy, number of sets of chromosomes. SBS, single base substitution signature, proportion of each substitution signature present per tumour. RS, rearrangement signature, proportion of each rearrangement signature per tumour. CN-LOH, copy neutral-loss of heterozygosity.

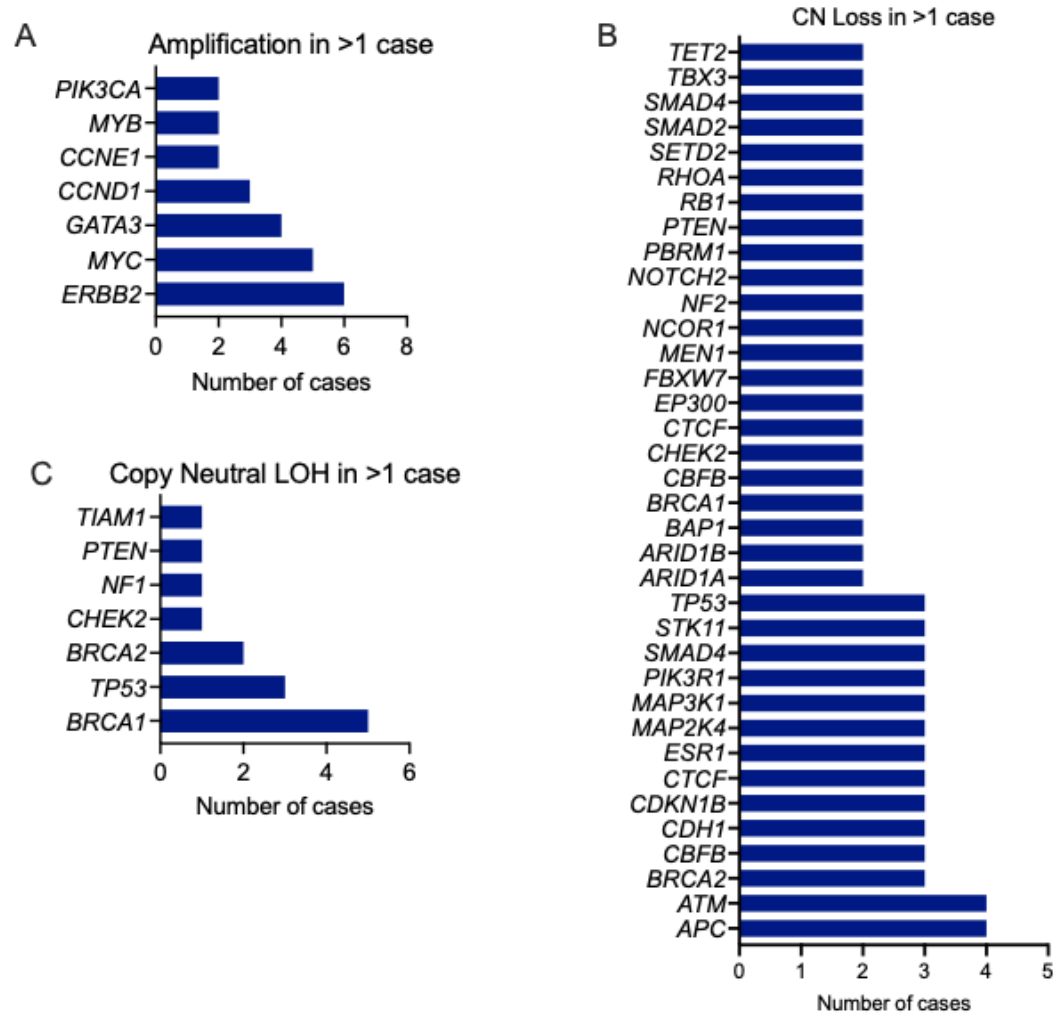


Figure 4: Somatic mutations and Copy number alterations in the Q-IMPROvE cohort. (A) the most frequently mutated genes are plotted, showing the long tail of genes mutated at low frequency. (B) *TP53* mutations were spread across the gene, as expected, while (C) the *PIK3CA* hotspot at amino acid 1047 was highly represented. Recurrent (D) single copy losses, (E) copy neutral Loss of Heterozygosity (CN LOH) and (F) amplifications of well-characterised oncogenes and tumour suppressor genes. (G) Correlation of *ERBB2* copy number from genome sequencing with HER2 Immunohistochemistry from pathology report; amplification is considered to be >6 copies. IHC, immunohistochemistry.

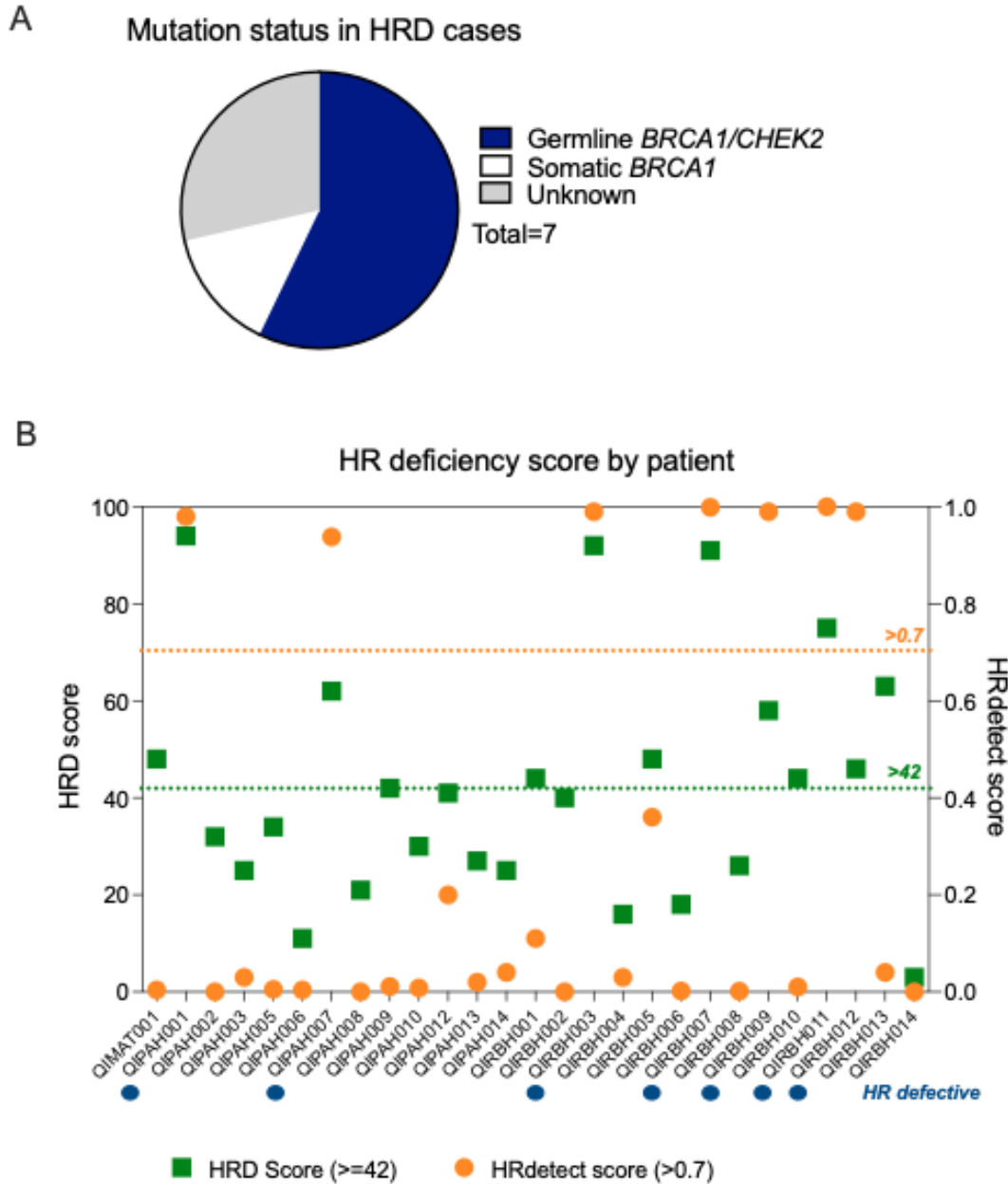


Figure 5: Homologous Recombination Deficiency and other actionable findings across the Q-IMPROVe cohort. (A) the seven HRD cases shown by likely cause. (B) All patients as a function of HRD Sum score (left Y axis) and HRDetect (right Y axis). The blue line indicates, the HRdetect cut-off score (0.7), while the grey line shows the HRD score cut-off (42). Patients marked with a purple circle meet requirements for both tests and are HR defective. (C) summary of actionable changes across the cohort. (D) the relationship between actionable changes and the outcome of neo-adjuvant therapy. No significant enrichment was detected ($P=0.1295$). pCR, pathologic complete response.

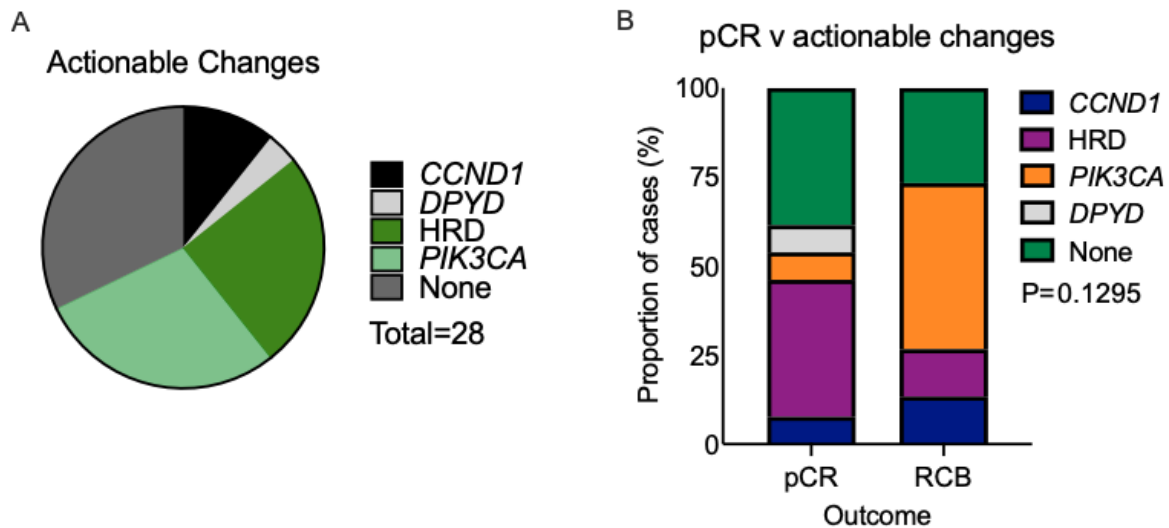


Figure 6: Actionable alterations identified in the Q-IMPROvE study. (A) summary of actionable changes across the cohort. (B) the relationship between actionable changes and the outcome of neo-adjuvant therapy. No significant enrichment was detected (P=0.1295).

Table 1 Breast cancer variants of interest reviewed in the Q-IMPROvE study

ENSEMBL Gene ID	Gene Symbol	Somatic/Germline
ENSG00000132142	ACACA	Somatic
ENSG00000184009	ACTG1	Somatic
ENSG00000077080	ACTL6B	Somatic
ENSG00000138071	ACTR2	Somatic
ENSG00000115091	ACTR3	Somatic
ENSG00000049192	ADAMTS6	Somatic
ENSG00000078295	ADCY2	Somatic
ENSG00000173020	ADRBK1	Somatic
ENSG00000155966	AFF2	Somatic
ENSG00000180772	AGTR2	Somatic
ENSG00000142208	AKT1	Somatic
ENSG00000204673	AKT1S1	Somatic
ENSG00000105221	AKT2	Somatic
ENSG00000161203	AP2M1	Somatic
ENSG00000134982	APC	Somatic
ENSG00000184945	AQP12A	Somatic
ENSG00000143761	ARF1	Somatic
ENSG00000160007	ARHGAP35	Somatic
ENSG00000141522	ARHGDI A	Somatic
ENSG00000117713	ARID1A	Somatic
ENSG00000049618	ARID1B	Somatic
ENSG00000111229	ARPC3	Somatic
ENSG00000171456	ASXL1	Somatic
ENSG00000156802	ATAD2	Somatic
ENSG00000123268	ATF1	Somatic
ENSG00000149311	ATM	Somatic and Germline (only variant ATM c.7271T>G)
ENSG00000175054	ATR	Somatic
ENSG00000085224	ATRX	Somatic
ENSG00000103126	AXIN1	Somatic
ENSG00000163930	BAP1	Somatic
ENSG00000183337	BCOR	Somatic
ENSG00000157764	BRAF	Somatic
ENSG00000012048	BRCA1	Somatic and Germline
ENSG00000139618	BRCA2	Somatic and Germline
ENSG00000156970	BUB1B	Somatic
ENSG00000135932	CAB39	Somatic
ENSG00000102547	CAB39L	Somatic
ENSG00000179218	CALR	Somatic
ENSG00000152495	CAMK4	Somatic
ENSG00000142453	CARM1	Somatic
ENSG00000064012	CASP8	Somatic
ENSG00000130940	CASZ1	Somatic
ENSG00000067955	CBFB	Somatic
ENSG00000114423	CBLB	Somatic
ENSG00000110092	CCND1	Somatic

ENSG00000112576	CCND3	Somatic
ENSG00000105173	CCNE1	Somatic
ENSG00000118816	CCNI	Somatic
ENSG00000117877	CD3EAP	Somatic
ENSG00000039068	CDH1	Somatic
ENSG00000170312	CDK1	Somatic
ENSG00000123374	CDK2	Somatic
ENSG00000135446	CDK4	Somatic
ENSG00000105810	CDK6	Somatic
ENSG00000124762	CDKN1A	Somatic
ENSG00000111276	CDKN1B	Somatic
ENSG00000147889	CDKN2A	Somatic
ENSG00000147883	CDKN2B	Somatic
ENSG00000172757	CFL1	Somatic
ENSG00000183765	CHEK2	Somatic and Germline (truncating only)
ENSG00000079432	CIC	Somatic
ENSG00000141367	CLTC	Somatic
ENSG00000088038	CNOT3	Somatic
ENSG00000005339	CREBBP	Somatic
ENSG00000204435	CSNK2B	Somatic
ENSG00000102974	CTCF	Somatic
ENSG00000044115	CTNNA1	Somatic
ENSG00000163131	CTSS	Somatic
ENSG00000257923	CUX1	Somatic
ENSG00000121966	CXCR4	Somatic
ENSG00000070190	DAPP1	Somatic
ENSG00000129187	DCTD	Somatic
ENSG00000175197	DDIT3	Somatic
ENSG00000119772	DNMT3A	Somatic
ENSG00000188641	DPYD	Somatic
ENSG00000108861	DUSP3	Somatic
ENSG00000101412	E2F1	Somatic
ENSG00000130159	ECSIT	Somatic
ENSG00000203734	ECT2L	Somatic
ENSG00000146648	EGFR	Somatic
ENSG00000151247	EIF4E	Somatic
ENSG00000100393	EP300	Somatic
ENSG00000141736	ERBB2	Somatic
ENSG00000065361	ERBB3	Somatic
ENSG00000178568	ERBB4	Somatic
ENSG00000175595	ERCC4	Somatic
ENSG00000091831	ESR1	Somatic
ENSG00000117560	FASLG	Somatic
ENSG00000109670	FBXW7	Somatic
ENSG00000158815	FGF17	Somatic
ENSG00000070388	FGF22	Somatic
ENSG00000111241	FGF6	Somatic
ENSG00000077782	FGFR1	Somatic
ENSG00000066468	FGFR2	Somatic

ENSG00000129514	FOXA1	Somatic
ENSG00000118689	FOXO3	Somatic
ENSG00000114861	FOXP1	Somatic
ENSG00000170324	FRMPD2	Somatic
ENSG00000128242	GAL3ST1	Somatic
ENSG00000107485	GATA3	Somatic
ENSG00000156049	GNA14	Somatic
ENSG00000087460	GNAS	Somatic
ENSG00000127928	GNGT1	Somatic
ENSG00000204175	GPRIN2	Somatic
ENSG00000132522	GPS2	Somatic
ENSG00000177885	GRB2	Somatic
ENSG00000120251	GRIA2	Somatic
ENSG00000155974	GRIP1	Somatic
ENSG00000082701	GSK3B	Somatic
ENSG00000196126	HLA-DRB1	Somatic
ENSG00000174775	HRAS	Somatic
ENSG00000197915	HRNR	Somatic
ENSG00000166598	HSP90B1	Somatic
ENSG00000140443	IGF1R	Somatic
ENSG00000147168	IL2RG	Somatic
ENSG00000113520	IL4	Somatic
ENSG00000136244	IL6	Somatic
ENSG00000160712	IL6R	Somatic
ENSG00000134352	IL6ST	Somatic
ENSG00000198001	IRAK4	Somatic
ENSG00000133124	IRS4	Somatic
ENSG00000123104	ITPR2	Somatic
ENSG00000143603	KCNN3	Somatic
ENSG00000147050	KDM6A	Somatic
ENSG00000055609	KMT2C	Somatic
ENSG00000167548	KMT2D	Somatic
ENSG00000133703	KRAS	Somatic
ENSG00000182866	LCK	Somatic
ENSG00000100097	LGALS1	Somatic
ENSG00000198799	LRIG2	Somatic
ENSG00000169032	MAP2K1	Somatic
ENSG00000034152	MAP2K3	Somatic
ENSG00000065559	MAP2K4	Somatic
ENSG00000108984	MAP2K6	Somatic
ENSG00000095015	MAP3K1	Somatic
ENSG00000135341	MAP3K7	Somatic
ENSG00000100030	MAPK1	Somatic
ENSG00000109339	MAPK10	Somatic
ENSG00000102882	MAPK3	Somatic
ENSG00000107643	MAPK8	Somatic
ENSG00000050748	MAPK9	Somatic
ENSG00000119487	MAPKAP1	Somatic
ENSG00000135679	MDM2	Somatic
ENSG00000112282	MED23	Somatic

ENSG00000133895	MEN1	Somatic
ENSG00000079277	MKNK1	Somatic
ENSG00000099875	MKNK2	Somatic
ENSG00000076242	MLH1	Somatic
ENSG00000130396	MLLT4	Somatic
ENSG00000095002	MSH2	Somatic
ENSG00000173531	MST1	Somatic
ENSG00000118513	MYB	Somatic
ENSG00000136997	MYC	Somatic
ENSG00000172936	MYD88	Somatic
ENSG00000008130	NADK	Somatic
ENSG00000158747	NBL1	Somatic
ENSG00000158092	NCK1	Somatic
ENSG00000124151	NCOA3	Somatic
ENSG00000141027	NCOR1	Somatic
ENSG00000196712	NF1	Somatic
ENSG00000186575	NF2	Somatic
ENSG00000104825	NFKBIB	Somatic
ENSG00000134259	NGF	Somatic
ENSG00000106100	NOD1	Somatic
ENSG00000148400	NOTCH1	Somatic
ENSG00000134250	NOTCH2	Somatic
ENSG00000213281	NRAS	Somatic
ENSG00000166368	OR2D2	Somatic
ENSG00000179468	OR9A2	Somatic
ENSG00000130669	PAK4	Somatic
ENSG00000083093	PALB2	Somatic and Germline (truncating only)
ENSG00000102699	PARP4	Somatic
ENSG00000075891	PAX2	Somatic
ENSG00000163939	PBRM1	Somatic
ENSG00000134853	PDGFRA	Somatic
ENSG00000152256	PDK1	Somatic
ENSG00000108518	PFN1	Somatic
ENSG00000156531	PHF6	Somatic
ENSG00000078142	PIK3C3	Somatic
ENSG00000121879	PIK3CA	Somatic
ENSG00000105851	PIK3CG	Somatic
ENSG00000145675	PIK3R1	Somatic
ENSG00000117461	PIK3R3	Somatic
ENSG00000115020	PIKFYVE	Somatic
ENSG00000127445	PIN1	Somatic
ENSG00000164093	PITX2	Somatic
ENSG00000123739	PLA2G12A	Somatic
ENSG00000182621	PLCB1	Somatic
ENSG00000124181	PLCG1	Somatic
ENSG00000122512	PMS2	Somatic
ENSG00000172531	PPP1CA	Somatic
ENSG00000137713	PPP2R1B	Somatic
ENSG00000057657	PRDM1	Somatic

ENSG00000046889	PREX2	Somatic
ENSG00000162409	PRKAA2	Somatic
ENSG00000181929	PRKAG1	Somatic
ENSG00000114302	PRKAR2A	Somatic
ENSG00000166501	PRKCB	Somatic
ENSG00000171862	PTEN	Somatic
ENSG00000179295	PTPN11	Somatic
ENSG00000134242	PTPN22	Somatic
ENSG00000153707	PTPRD	Somatic
ENSG00000188060	RAB42	Somatic
ENSG00000136238	RAC1	Somatic
ENSG00000132155	RAF1	Somatic
ENSG00000144118	RALB	Somatic
ENSG00000139687	RB1	Somatic
ENSG00000147274	RBMX	Somatic
ENSG00000132677	RHBG	Somatic
ENSG00000067560	RHOA	Somatic
ENSG00000137275	RIPK1	Somatic
ENSG00000143622	RIT1	Somatic
ENSG00000156313	RPGR	Somatic
ENSG00000117676	RPS6KA1	Somatic
ENSG00000177189	RPS6KA3	Somatic
ENSG00000141564	RPTOR	Somatic
ENSG00000159216	RUNX1	Somatic
ENSG00000181555	SETD2	Somatic
ENSG00000115524	SF3B1	Somatic
ENSG00000175793	SFN	Somatic
ENSG00000160691	SHC1	Somatic
ENSG00000155926	SLA	Somatic
ENSG00000117394	SLC2A1	Somatic
ENSG00000175387	SMAD2	Somatic
ENSG00000141646	SMAD4	Somatic
ENSG00000127616	SMARCA4	Somatic
ENSG00000115904	SOS1	Somatic
ENSG00000065526	SPEN	Somatic
ENSG00000161011	SQSTM1	Somatic
ENSG00000101972	STAG2	Somatic
ENSG00000170581	STAT2	Somatic
ENSG00000168610	STAT3	Somatic
ENSG00000118046	STK11	Somatic
ENSG00000183735	TBK1	Somatic
ENSG00000177565	TBL1XR1	Somatic
ENSG00000135111	TBX3	Somatic
ENSG00000121075	TBX4	Somatic
ENSG00000089225	TBX5	Somatic
ENSG00000124678	TCP11	Somatic
ENSG00000168769	TET2	Somatic
ENSG00000159445	THEM4	Somatic
ENSG00000156299	TIAM1	Somatic
ENSG00000244045	TMEM199	Somatic

ENSG0000067182	TNFRSF1A	Somatic
ENSG00000141510	TP53	Somatic and Germline
ENSG00000127191	TRAF2	Somatic
ENSG00000101255	TRIB3	Somatic
ENSG00000116747	TROVE2	Somatic
ENSG00000103197	TSC2	Somatic
ENSG00000109332	UBE2D3	Somatic
ENSG00000177889	UBE2N	Somatic
ENSG00000124486	USP9X	Somatic
ENSG00000134215	VAV3	Somatic
ENSG00000100219	XBP1	Somatic
ENSG00000166913	YWHAB	Somatic
ENSG00000185650	ZFP36L1	Somatic
ENSG00000147130	ZMYM3	Somatic
ENSG00000171940	ZNF217	Somatic
ENSG00000198538	ZNF28	Somatic
ENSG00000160094	ZNF362	Somatic
ENSG00000183779	ZNF703	Somatic

Table 2 Pharmacogenomic variants of interest

Gene	Drug	Response	RS number
<i>BRINP1</i>	Trastuzumab	cardiotoxicity (decline in LVEF)	rs10117876
<i>C10orf11</i>	Tamoxifen	Recurrence-free survival	rs10509373
<i>DPYD</i>	Fluoropyrimidines	Severe toxicity	rs55886062
<i>EPHA5</i>	Paclitaxel	Sensory neuropathy	rs7349683
<i>FGD4</i>	Paclitaxel	Sensory neuropathy	rs10771973
intergenic_chr14 (<i>TCL1A</i>)	Anastrozole, exemestane	Musculoskeletal adverse events	rs11849538
Intergenic_region_of_chr6p22.3	Trastuzumab	cardiotoxicity (decline in LVEF)	rs4305714
<i>LDB2</i>	Trastuzumab	Cardiotoxicity (decline in LVEF)	rs55756123
<i>RAB22A</i>	Trastuzumab	cardiotoxicity (decline in LVEF)	rs70755
<i>SPRR1A/CACNB4</i>	Combinations of chemotherapy	Alopecia	rs3820706
<i>SV2C</i>	Bevacizumab	Hypertension	rs6453204
<i>TPD52</i>	Lapatinib hepatotoxicity	Lapatinib hepatotoxicity	rs7828135
<i>TRPC6</i>	Trastuzumab	cardiotoxicity (decline in LVEF)	rs77679196

ZNF613	Endocrine therapy	Survival	rs8113308
intergenicVariant_chr15	Anthracycline	Congestive heart failure	rs28714259
regulatoryRegionVariant_chr8/ Intergenic region of chr8q21.11	Anastrozole, exemestane	Breast cancer-free interval	rs13260300

Table 3 Sequencing Quality Control Data

QIMPROVE_ID	Genomiqa_ID	Bam_name	Sample_type	Q30%	Unmapped_reads%	Duplicate_reads%	Average read depth
QI_1	GQ21P000024	8f3c8dea-77a4-4257-8413-30cd804d1763.bam	Normal	81.19	0.7	3.72	33.6
QI_1	GQ21P000024	529b3f50-5dd7-4124-8b8b-a0b99152b0ae.bam	Tumour	79.79	0.53	3.75	70.8
QI_2	GQ21P000007	937a30e0-6c65-4ffa-814b-8dd1d8a2e106.bam	Normal	85.09	0.17	2.91	36.9
QI_2	GQ21P000007	f2bfb8f2-e5cc-4cf-d-ab36-88f5faca194c.bam	Tumour	84.44	0.19	4.33	74.9
QI_3	GQ21P000008	a27feb2d-6f52-4248-af7f-8dcace6cb430.bam	Normal	89.81	0.25	3.18	40.6
QI_3	GQ21P000008	1664cfb5-b432-4ecb-9dcd-04d250d54b29.bam	Tumour	92.72	0.09	2.8	80
QI_4	GQ21P000009	fc468cba-3c00-4e67-b4eb-bda01970ffdb.bam	Normal	85.05	0.21	2.49	37.6

QI_4	GQ21P000009	e4665511-031a-491e-86de-0e94b89fe058.bam	Tumour	85.92	0.2	3.96	69.7
QI_6	GQ21P000028	a90ae85d-307a-47fb-9d7b-dace2d3288a6.bam	Normal	80.47	0.5	2.68	35.9
QI_6	GQ21P000028	15ea8134-9f74-4424-b2d7-e5fb637ac5dd.bam	Tumour	80.9	0.51	3.78	73.2
QI_7	GQ21P000023	461c666f-0692-4abe-851b-70e80438ce31.bam	Normal	83.01	0.43	2.82	36.7
QI_7	GQ21P000023	c8d7e81d-52f5-4e9a-8460-180172158531.bam	Tumour	77.15	0.66	4.83	90.3
QI_8	GQ21P000029	0b3faeb7-e51a-48dc-b112-a57c3d6b1eb9.bam	Normal	82.46	0.39	2.61	39.5
QI_8	GQ21P000029	b204f0beced8-4247-887b-3caeabd9d761.bam	Tumour	78.03	0.63	5.26	55.8
QI_9	GQ21P000027	ae4bf69f-c002-4c0b-9860-f8bd54fe4372.bam	Normal	80.91	0.55	2.65	34.6
QI_9	GQ21P000027	8afc852f-895b-4a10-a12e-85b666e81778.bam	Tumour	78.23	0.59	4.26	56.7
QI_10	GQ21P000022	dc8e46b9-19c3-40eb-aa36-b76fc36a7705.bam	Normal	80.28	0.48	2.74	36

QI_10	GQ21P000 022	13e98439- 5306-4ac7- 810a- 739ce65d302 6.bam	Tumou r	80.89	0.52	4.14	71.4
QI_11	GQ21P000 020	e93653a8- 904c-4160- a1c8- c265ae10e078 .bam	Normal	81.83	0.4	2.57	38.1
QI_11	GQ21P000 020	803ba786- d253-43b3- 832c- f33416c8f902. bam	Tumou r	80.51	0.47	3.43	68.6
QI_12	GQ21P000 025	359d156d- 5e0e-4263- 9d2e- a939c9461d4 a.bam	Normal	79.99	0.48	2.42	35.7
QI_12	GQ21P000 025	b7e8080e- 8f16-44fe- b83b- 66974f24c318 .bam	Tumou r	78.68	0.55	3.93	67.1
QI_13	GQ21P000 021	fc266455- 192c-441b- 95fe- a29e2b96f571 .bam	Normal	79.84	0.48	2.04	32.2
QI_13	GQ21P000 021	a4b2275b- 5090-4716- 9972- b11c6c52f575 .bam	Tumou r	80.77	0.49	3.13	74.5
QI_14	GQ21P000 026	32bff9ae- 03cb-4a1b- 9fab- 29b0bd4bd5e 5.bam	Normal	78.93	0.49	2.57	31.2
QI_14	GQ21P000 026	92f2938b- 1a14-4815- a502- d16e9583bcf8 .bam	Tumou r	79.76	0.52	2.98	74.8
QI_15	GQ21P000 039	94762fa4- 3637-4f76- b556- 2a7b99d0b3e 3.bam	Normal	85.69	0.27	2.65	38.7

QI_15	GQ21P000039	9b88c6d4-3960-4b08-b2dc-27eeb547faf9.bam	Tumour	86.61	0.24	3.29	79.7
QI_16	GQ20P000001	1a295f6b-0acd-4c65-955c-d2ec8450cdf.bam	Normal	83.13	0.21	2.56	35.4
QI_16	GQ20P000001	02350dea-f256-40ed-aded-9d756fc8141d.bam	Tumour	84.92	0.19	4.1	70.7
QI_17	GQ20P000002	ec8079ab-38d4-4000-8279-fc17ed7fe4cb.bam	Normal	82.43	0.22	2.77	33.6
QI_17	GQ20P000002	b5a14cca-07fc-4df1-a5a9-56e0e0753bb2.bam	Tumour	84.56	0.2	4.11	74.7
QI_18	GQ20P000003	46aa8920-2fab-47c4-9bad-55f901b4ae91.bam	Normal	83.39	0.19	3.03	36.7
QI_18	GQ20P000003	045a53bc-dd6d-41de-aecc-54dce5c5925d.bam	Tumour	84.61	0.21	4.09	70.9
QI_19	GQ21P000001	741dfade-0103-4f1e-a808-545e3793bf4e.bam	Normal	84.06	0.19	2.98	36.8
QI_19	GQ21P000001	34ff3f04-7dab-4435-8678-eead1a3e7475.bam	Tumour	86.5	0.16	4.5	74.2
QI_20	GQ21P000002	b884cd7f-06f0-4740-947e-2ee140e669c8.bam	Normal	86.24	0.2	2.74	37.2

QI_20	GQ21P000002	3a6041f7-a631-4a45-966d-69d2c7bef42a.bam	Tumour	85.63	0.22	4.94	67.7
QI_21	GQ21P000003	f1191fe4-33b7-4b36-8676-23dd99cb818c.bam	Normal	84.99	0.22	2.97	34.5
QI_21	GQ21P000003	d7b00831-a940-42f2-a5c3-6d9770a3ee22.bam	Tumour	85.47	0.21	5.24	77.4
QI_22	GQ21P000004	8143038e-43c0-44f4-9bf2-fb4731b96e51.bam	Normal	86.43	0.19	3.17	37.4
QI_22	GQ21P000004	d10747b1-2b15-49d3-a402-3a44712f4b47.bam	Tumour	85.06	0.22	3.57	66
QI_23	GQ21P000005	1fc98c9c-779f-4dae-bea3-d46d18a90d9d.bam	Normal	85.97	0.22	2.61	36.2
QI_23	GQ21P000005	0edb8d4c-2f22-40b5-a5a7-56832b3e700d.bam	Tumour	86.87	0.23	3.7	68.7
QI_24	GQ21P000006	d484d3b0-7920-49b8-b4fc-0a67b2cc98b7.bam	Normal	91.5	0.08	2.57	37.5
QI_24	GQ21P000006	1d0050ae-40e9-4a7f-be5f-b5ae16c2d2b9.bam	Tumour	87.42	0.39	3.24	77.1
QI_25	GQ21P000036	841df1d8-aeae-4e75-a37a-368d27960c8c.bam	Normal	86.73	0.23	2.58	39.1

QI_25	GQ21P000036	74a34589-3e03-4f98-8f26-66f3b15f90a9.bam	Tumour	86.51	0.24	3.41	71.2
QI_26	GQ21P000031	06fee153-d1c9-49de-9b80-313bb4b8af8c.bam	Normal	86.7	0.21	2.36	36.4
QI_26	GQ21P000031	54701c94-6dbb-49ca-a606-8fc8a689eb00.bam	Tumour	91.81	0.09	3.9	78.5
QI_27	GQ21P000032	5f32db8f-9f9e-4e54-8675-638c95902334.bam	Normal	87.02	0.21	2.3	39.1
QI_27	GQ21P000032	df89efc9-38b1-4ee6-a78b-9bf1533e2042.bam	Tumour	91.63	0.07	3.77	79.7
QI_28	GQ21P000037	1d320741-7591-4957-a320-17bacd761781.bam	Normal	86.74	0.23	2.49	37.3
QI_28	GQ21P000037	2134047c-add3-4181-bb3c-aa3d3e9750a7.bam	Tumour	92.29	0.08	4.22	80
QI_29	GQ21P000038	44a801f5-179b-4970-bc29-2b9054bc698a.bam	Normal	85.48	0.32	2.4	38.2
QI_29	GQ21P000038	94ee75b7-cd63-4e21-b577-403f8332d96c.bam	Tumour	92.52	0.07	3.96	80.7

Table 4 Breast cancer relevant germline findings in the Q-IMPROvE cohort

Age [^]	Existing Genetic Health referral for BrCa panel*	Germline Status at recruitment	WGS germline finding	Putative somatic second hit	WGS Somatic finding of interest	New referral to Genetic Health
25-35	Y	None	None	Copy neutral LOH <i>TP53</i>	<i>TP53</i> p.Cys135Tyr	Yes
25-35	Y	<i>BRCA1</i> mutant	<i>BRCA1</i> p.Asn1355Lysfs Ter10		.	.
25-35	Y	None	None		.	.
45-55	Y	None	None		.	.
45-55	Y	None	None		.	.
45-55	Y	<i>BRCA1</i> mutant	<i>BRCA1</i> p.Gly1348Asnfs Ter7	Copy loss <i>BRCA2</i>	.	.
45-55	Y	None	None		.	.
45-55	N	None	<i>BRCA1</i> p.Val627SerfsTer4		.	Yes
45-55	N	None	<i>CHEK2</i> p.Gln20Ter	Copy neutral LOH <i>CHEK2</i>	.	Yes
60+	Y	None	None	Copy neutral LOH <i>BRCA1</i>	.	.

[^]patient fits within age bracket indicated;

*local clinical, germline breast cancer panel comprises: ATM (only variant c.7271T>G); *BRCA1*; *BRCA2*; *CHEK2* (truncating only); *PALB2* (truncating only); *TP53*.

BrCa, breast cancer; LOH, Loss of heterozygosity; WGS, Whole genome sequencing.

Table 5 Frequency of copy number alterations in the Q-IMPROvE cohort

Gene	# High Gain	Gene	# Copy Loss	Gene*	# CN LOH	Gene	# HOMD
<i>ERBB2</i>	6	<i>APC</i>	4	<i>BRCA1</i>	5	<i>CDKN2A</i>	1
<i>MYC</i>	5	<i>ATM</i>	4	<i>TP53</i>	3	<i>CDKN2B</i>	1
<i>GATA3</i>	4	<i>BRCA2</i>	3	<i>BRCA2</i>	2		
<i>CCND1</i>	3	<i>CBFB</i>	3	<i>CHEK2</i>	1		
<i>CCNE1</i>	2	<i>CDH1</i>	3	<i>NF1</i>	1		
<i>MYB</i>	2	<i>CDKN1B</i>	3	<i>PTEN</i>	1		
<i>PIK3CA</i>	2	<i>CTCF</i>	3	<i>TIAM1</i>	1		
<i>AKT1</i>	1	<i>ESR1</i>	3				

<i>ATAD2</i>	1	<i>MAP2K4</i>	3				
<i>BRAF</i>	1	<i>MAP3K1</i>	3				
<i>FOXO3</i>	1	<i>PIK3R1</i>	3				
<i>GNAS</i>	1	<i>SMAD4</i>	3				
<i>PREX2</i>	1	<i>STK11</i>	3				
<i>RAF1</i>	1	<i>TP53</i>	3				
<i>SLA</i>	1	<i>ARID1A</i>	2				
<i>TBL1XR1</i>	1	<i>ARID1B</i>	2				
<i>TBX4</i>	1	<i>BAP1</i>	2				
<i>TP53</i>	1	<i>BRCA1</i>	2				
		<i>CBFB</i>	2				
		<i>CHEK2</i>	2				
		<i>CTCF</i>	2				
		<i>EP300</i>	2				
		<i>FBXW7</i>	2				
		<i>MEN1</i>	2				
		<i>NCOR1</i>	2				
		<i>NF2</i>	2				
		<i>NOTCH2</i>	2				
		<i>PBRM1</i>	2				
		<i>PTEN</i>	2				
		<i>RB1</i>	2				
		<i>RHOA</i>	2				
		<i>SETD2</i>	2				
		<i>SMAD2</i>	2				
		<i>SMAD4</i>	2				
		<i>TBX3</i>	2				
		<i>TET2</i>	2				
		<i>ATRX</i>	1				
		<i>AXIN1</i>	1				
		Gene	# Copy Loss				
		<i>BAP1</i>	1				
		<i>BCOR</i>	1				
		<i>BRCA1</i>	1				
		<i>CASP8</i>	1				
		<i>CDH1</i>	1				
		<i>CDKN2A</i>	1				
		<i>CHEK2</i>	1				
		<i>CREBBP</i>	1				
		<i>EP300</i>	1				

		<i>ERBB4</i>	1				
		<i>ERCC4</i>	1				
		<i>FOXO3</i>	1				
		<i>KDM6A</i>	1				
		<i>KMT2C</i>	1				
		<i>KMT2D</i>	1				
		<i>MAP2K4</i>	1				
		<i>MAP3K1</i>	1				
		<i>MLH1</i>	1				
		<i>NCOR1</i>	1				
		<i>NF1</i>	1				
		<i>NF2</i>	1				
		<i>PALB2</i>	1				
		<i>PBRM1</i>	1				
		<i>PHF6</i>	1				
		<i>PMS</i>	1				
		<i>PRDM1</i>	1				
		<i>PTPRD</i>	1				
		<i>RHOA</i>	1				
		<i>SETD2</i>	1				
		<i>SMAD2</i>	1				
		<i>SMARCA4</i>	1				
		<i>STAG2</i>	1				
		<i>TSC2</i>	1				
* CN LOH reported only for those genes with a mutation also recorded							
CN, copy neutral; HOMD, homozygous deletion; LOH, Loss of heterozygosity.							