Whole genome sequencing reveals differences in the genetic landscape of interval and screen-detected breast cancer

Charlie Mills*[1], Amit Sud*[1,2,3,4,5], Andrew Everall*[1], Dan Chubb[1], Samuel E. D. Lawrence[1], Ben Kinnersley[1,6], Alex J. Cornish[1], Robert Bentham[6], Richard S. Houlston[1]

1. Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK
2. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA.
3. Harvard Medical School, Boston, MA, USA.
4. Broad Institute of MIT and Harvard, Cambridge, MA, USA.
5. Centre of Immuno-Oncology, Nuffield Department of Medicine, University of Oxford, Oxford, UK.
6. University College London Cancer Institute, University College London, London, UK.

* These authors contributed equally to this work.

Correspondence to: Richard Houlston; +44(0) 208 722 4175; richard.houlston@icr.ac.uk

Running title: Genetic landscape of interval and screen-detected breast cancer

Number of words: 3460

ABSTRACT

Interval breast cancers (IBCs) are cancers diagnosed between screening episodes. Understanding the biological differences between IBCs and screen-detected breast-cancers (SDBCs) has the potential to improve mammographic screening and patient management. We analysed and compared the genomic landscape of 288 IBCs and 473 SDBCs by whole genome sequencing of paired tumour-normal patient samples collected as part of the UK 100,000 Genomes Project. Compared to SDBCs, IBCs were more likely to be lobular, higher grade, and triple negative. A more aggressive clinical phenotype was reflected in IBCs displaying features of genomic instability including a higher mutation rate and number of chromosomal structural abnormalities, defective homologous recombination and *TP53* mutations. We did not however, find evidence to indicate that IBCs are associated with a significantly different immune response. While IBCs do not represent a unique molecular class of invasive breast cancer they exhibit a more aggressive phenotype, which is likely to be a consequence of the timing of tumour initiation. This information is relevant both with respect to treatment as well as informing the screening interval for mammography.

INTRODUCTION

The high survival rates associated with the early detection of localised breast cancer together with the screening properties of mammography, have provided the rationale for population breast cancer screening programs[1–3]. In the UK, the National Health Service (NHS) invites women aged between 50 and 70 for 3-yearly mammography, but with provision for those aged 45 and over to be screened if self-referred. Although studies suggest a mammography every 2-3 years may reduce breast cancer-specific mortality by a fifth[4] around 30% of breast cancers in women attending screening are interval breast cancers (IBCs) diagnosed between screening episodes[5].

While technical and patient factors may in part explain the incidence of IBCs[6,7], studies have suggested IBCs may represent a more aggressive form of breast cancer[8]. This would accord with the observation that most IBCs are not visible on the index screen in retrospective review[8]. Exploring possible genetic differences between IBCs and screen-detected breast cancers (SDBCs) has the potential to inform mammographic screening and patient management. Previous analyses of IBCs have so far predominantly focused on specific genes and studies have typically been based on small patient numbers[9,10].

To provide the most comprehensive analysis of the genomic landscape of IBC, we analysed whole genome sequencing (WGS) data generated on 288 IBCs recruited to the UK's 100,000 Genomes Project (100kGP). The results of our study highlight distinct clinical, genetic and evolutionary differences between IBC and SDBC (**Fig. 1**).

RESULTS

Clinico-pathological differences

After curation we based our analysis on 288 IBCs and 473 SDBCs. The clinical details of the 761 cases are summarised in **Table 1**. Overall, 72% (n=548) of the cases were self-reported to be White British and there was no difference in the frequency of ethnicities between IBCs and SDBCs. There was no significant difference in the tumour purity of IBCs and SDBCs (P = 0.17), making it unlikely that study findings will have been affected by sample collection and processing. Interval breast cancers were more likely to be larger tumours than SDBCs (median size 24mm and 20mm respectively, P = 1.9 x 10^{-5}), be lobular (OR = 1.61, 95% confidence interval (CI) 1.07-2.42, P = 0.02), be higher grade tumours (OR = 1.83, 95% CI 1.45-2.31, P = 3.8 x 10^{-7}), be oestrogen receptor (ER) negative (OR = 0.43, 95% CI 0.26-0.73, P = 1.11 x 10^{-3}), be progesterone receptor (PR) negative (OR = 0.40, 95% CI 0.25-0.65, P = 4.45 x 10^{-5}), be TN (OR = 2.01, 95% CI: 1.02-3.98, P = 0.03) and patients were more likely to be diagnosed with nodal disease at presentation (OR = 1.56, 95% CI: 1.12-2.17, P = 7.13 x 10^{-3}) (**Table 1**).

Patients with lobular IBCs were more likely to be older at diagnosis than those with SDBCs (mean ages 61.1 years and 58.5 years respectively; P = 0.03). While patients with ductal TN IBCs were older than those with SDBCs (mean age 61.1 years and 56.6 years respectively; P = 0.03), patients with ductal ER-positive breast IBCs cancers tended to be younger than those with SDBC with the same histology (mean ages 58.2 years and 59.8 years respectively; P = 0.02).

Germline variation and polygenic scores

We identified 17 patients with IBCs who were carriers of a pathogenic mutation in one of the HBOC genes (13 *BRCA2*, 2 *CHEK2*, 1 *MSH6* and 1 *BRCA1*) and 27 patients with SDBC who were HBOC gene carriers (12 *BRCA2,* 7 *ATM*, 3 *CHEK2*, 1 *BARD1*, 1 *BRIP1*, 1 *BRCA1*, 1 *MSH6*, 1 *PALB2*) (**Supplementary Table 2**). The frequency of patients with *BRCA*-mutations was marginally, but not statistically significantly, higher in IBCs compared to SBDCs (OR= 1.80, 95% 0.77-4.23, *P*=0.16). Considering all HBOC genes, there was also no significant difference in carrier frequency (5.9% versus 5.7%).

We generated PGS on a per sample basis to investigate the relationship between genetically predicted BMI and breast density as well as modifiable breast cancer risk factors in IBC and SDBC. In a case-only analysis, while SDBCs had a higher PGS for overall breast cancer than IBC, this was not significant (*P* = 0.68). Similarly, there was no difference in genetically predicted breast density or BMI between IBCs and SDBCs (*P*-values 0.09 and 0.65 respectively; **Supplementary Table 3**).

Somatic alterations

The frequency of somatic mutations (single nucleotide variants and indels) was significantly higher in all IBCs compared to all SDBCs (1.47/Mb (0.18 − 18.55) and 1.06/Mb (0.08 − 19.05) respectively, $P = 1.43 \times 10^{-3}$). This significant difference was also shown when stratified by tumour subtype - all ductal ($P = 6.32 \times 10^{-4}$), ductal ER-positive ($P = 5.23 \times 10^{-3}$) and ductal non-TN tumours ($P = 1.81 \times 10^{-3}$) (**Fig. 2, Supplementary Fig. 1**). After adjustment for tumour grade the mutational burden remained significantly higher in all IBCs compared to all SDBCs ($P = 0.03$).

In the 67 putative invasive breast cancer drivers identified in the 100kGP data, we identified 539 predicted oncogenic mutations in IBCs and 931 predicted oncogenic mutations in SDBCs[37]. As expected from previously published studies of breast cancer in both IBCs and SDBCs tumours[38] the most commonly mutated breast cancer driver genes included *PIK3CA, TP53, KMT2C, GATA3, CDH1, MAP3K1* and *PTEN* (**Supplementary Table 4**). Furthermore, mutational frequencies in drivers differed significantly between ductal and lobular breast cancer histologies irrespective of screen detection status (**Supplementary Fig. 2**). There were, however, significant differences in the frequency of mutations in these driver genes between all IBCs and all SDBCs and when the analysis was only confined to ductal tumours. (**Fig. 3**). For completeness we examined for differences in the frequency of mutations in all genes between IBC and SDBC, finding no further significant differences (**Supplementary Table 5**).

The frequency of *TP53* mutations was higher in IBCs compared with SDBCs (36% vs 24%, OR = 1.80, 95% CI: 1.29-2.51, P = 3.64 x 10^{-4}), while the frequency of PIK3CA (33% vs 45%, OR: 0.59, 95% CI: 0.43-0.80, P = 6.11 x 10^{-4}) and *MAP3K1* mutations was lower (IBC 8% vs SDBC 17%, OR = 0.43, 95% CI: 0.25-0.71, P = 4.34 x 10^{-4}) (**Fig. 3**). These differences in mutational frequency remained significant when the analysis was confined to ductal histology and non-TN breast cancers. No significant difference in the frequency of mutations in oncogenic driver genes was shown between lobular IBC and SDBC, likely a consequence of limited statistical power. When considering the clonality of driver mutations, only *TP53* mutations were significantly more likely to be clonal in IBCs (P = 0.04, OR = 3.43, 95% CI 1.02-14.96).

While the profile of somatic copy number alterations was broadly consistent between all IBCs and SDBCs (**Fig. 4**), the overall fraction of the genome altered was significantly higher in IBCs ($P = 2.56 \times 10^{-4}$). This difference remained significant when restricted to ductal ($P = 9 \times 10^{-6}$) and ER-positive ($P = 4.67 \times 10^{-3}$) subtypes (**Fig. 5, Supplementary Fig. 3**). Additionally, the number of CNAs was significantly higher in IBCs ($P = 9.72 \times 10^{-4}$) (**Supplementary Table 6**). Whole genome duplication (WGD) was also a significantly more frequent event for all ($P = 8.51 \times 10^{-3}$, OR = 1.59, 95% CI 1.11-2.27), ductal ($P = 3.14 \times 10^{-4}$, OR = 1.98, 95% CI 1.35-2.91) and non-TN IBCs ($P = 0.01$, OR = 1.76, 95% CI 1.11-2.79) compared to corresponding SDBCs. However, after adjusting for grade, the fraction of genome altered and number of CNAs were no longer significantly different between IBCs and SDBCs ($P = 0.27$, $P = 0.50$ respectively), reflecting the observation that IBCs are more likely to be higher grade.

Structural variants were a feature of almost all cancers (99% IBC vs 98% SDBC) and classification of SVs revealed that both balanced ($P = 5.87 \times 10^{-3}$, OR = 1.74, 95% CI 1.16-2.63) and unbalanced ($P = 8.79 \times 10^{-3}$, OR = 1.49, 95% CI 1.10-2.02) translocation events were enriched in IBCs. However, an adjustment for tumour grade greatly attenuated the effect such that only unbalanced translocations remained significantly enriched in IBCs ($P = 0.03$, OR = 1.03, 95% CI 1.00-1.06). No other patterns of structural rearrangements, including chromothripsis, were found to be statistically enriched or depleted in IBCs.

Mutational signatures

To examine if the mutational processes in IBC and SDBC are different we utilised signature activities that have been previously extracted *de novo* and related to known COSMIC

signatures[18] (**Supplementary Table 7, Supplementary Fig. 4**). Across all cancers the most common signatures were those resulting from clock-like mutagenic processes (SBS1 and SBS5). As expected, signatures indicative of dHR were strongly associated with TN breast cancers, regardless of screening status. Across both IBC and SDBC the aetiological basis of dHR was only identifiable in 13% of cases on the basis of biallelic inactivation of *BRCA1, BRCA2, PALB2, BRIP1* or *RAD51B* through germline mutations, somatic mutations and loss of heterozygosity. A further 83% of cases harboured a monoallelic loss in these genes and hence many of the remaining cases may be caused by promoter methylation; however, this data is not available for 100kGP samples. Stratifying by screening status, dHR was significantly more common in IBC (OR = 2.55, 95% CI: 1.33-4.98, P = 3.38 x 10^{-3}) and this association was primarily driven by ductal ER+ tumours (OR = 4.90, 95% CI: 1.79-14.81, P = 6.95 x 10^{-4}). Moreover, signatures associated with dHR had significantly higher activities on average in IBC compared to SDBC: SBS3 (P = 0.02), ID6 (P = 1.79 x 10^{-4}) and CN17 (P = 2.65 x 10^{-3}). However, when restricting to TN breast cancers, dHR was not significantly enriched or depleted in IBCs (P = 0.66).

Immune evasion

We predicted the presence of 12,785 neoantigens across both IBCs and SDBCs. There was no significant difference in neoantigen burden between IBCs and SDBCs (P = 0.26). The frequency of somatic mutations and LOH of HLA class I genes was not significantly different between IBCs and SDBCs. There was also no evidence to support preferential inactivation of an APG in IBC. Finally, there was no significant difference in the TCRA T-cell fraction between IBCs and SDBCs (P = 0.83).

DISCUSSION

In the 100kGP cohort, IBCs were characterised by more aggressive tumour characteristics than SDBCs - a higher histological grade, larger tumour size, more frequently lymph node-positive at diagnosis, a higher proportion of ER and PR negativity, and more likely to be TN. These observations are consistent with several published studies[7,39–56]. While not statistically significant, but in keeping with a recent study, IBC patients were more likely to be carriers of BRCA-mutations than those with SDBC (4.5% vs 2.5%)[57].

Adjusting for grade, IBCs displayed a higher mutation rate while copy number abnormalities were not significantly different. IBCs additionally showed salient differences in driver mutational profile notably with a lower frequency of *PIK3CA* mutations and higher frequency of *TP53* mutations. These *TP53* mutations in IBCs were more likely to be clonal and hence by inference more likely to have arisen early in tumour development or confer a survival advantage to all tumour cells. This supports the notion that IBCs have differing evolutionary histories when compared to SDBCs[12,22]. The IBCs were also more likely to display dHR, which in part may reflect a higher rate of germline *BRCA* mutations. This is in keeping with published studies reporting a higher rate of germline HBOC predisposition in IBC compared to SDBC, as well as elevated rates of interval cancers in HBOC mutation carriers[9,57–63]. We did not, however, find evidence to indicate that women with IBC are at a lower PGS-defined breast cancer risk which has been reported by another analysis[64]. Similarly, while it has been proposed that IBC may be associated with a different immune response to SDBC[65], our analysis of genetically mediated immune evasion or T-cell tumour infiltration provides no support for such an assertion.

Our findings are consistent with recently published studies reporting higher rates of *TP53* mutations and dHR in IBC[9,10]. Cheasley *et al.*[9] also reported cancers in low mammographic density breasts (n=142) had a higher frequency of *TP53* mutations, dHR, higher fraction of the genome altered, more copy number gains and were more likely to be interval breast cancers when compared to cancers in high mammographic density breasts (n=119)[65].

The major strengths of our study are having access to a unique dataset with high-coverage WGS data, linkage to high-quality clinicopathological information features on patients screened in the context of population-wide screening, thus avoiding biases related to screening indication. Furthermore, rather than focusing on a restricted set of genes we have been able to undertake a systematic analysis of the genetic landscape of IBCs. This has allowed us to assess tumour intrinsic mechanisms contributing to IBC as well as surrogates for the tumour microenvironment (TME).

We do, however, appreciate there are a number of limitations to our study. It is the case that while lobular cancers tend not to be a rapidly growing tumour, they are more likely to be missed by screening because of a diffuse growth pattern and minimal stromal response[66]. Indeed, in our analysis the mutational rate in lobular IBC was only marginal higher than in lobular SDBC. Hence, we acknowledge that a subset of lobular IBCs might not necessarily be true interval cancers but ones that are missed by screening. While some other cases in the IBC group may also be missed because of high breast density, it is noteworthy that the mutational rate in ductal IBCs was far higher than in ductal SDBC, which provides support for a different biology rather than solely being ascribable to the consequence of screening performance. Moreover in the absence of measured breast density we have relied on PGS in

assessing the role of breast density and other modifiable factors, which inevitably affords limited power to demonstrate a relationship since these PGSs only capture a small proportion of the phenotypic variation of each risk factor[34,67]. Additionally, we did not have access to expression data and other classifiers of breast cancer histology such as the Gallen subtypes or PAM50[68,69]. Finally, we made use of indirect measurements of the TME and were unable to assess the impact of immune dysfunction in the earliest stages of tumour development.

As well as informing on the biology of IBCs, the findings of our study may inform on the detection and management of breast cancer. The higher number of *TP53* and copy number alterations as well as an increased mutation rate and dHR, suggests genomic instability has a greater role in IBCs when compared to SDBC. Genomic instability may be exploited therapeutically through synthetic lethality[70,71], such as sensitivity to PARP in the context of *BRCA*-deficiency[72,73]. Furthermore, given 30% of breast cancers in women regularly attending screening are IBCs, additional screening modalities have been advocated as a means to advance the early detection of breast cancer. Whilst early in development, noninvasive detection of chromosomal instability in plasma circulating cell-free DNA has shown promise and may serve as an adjunct to breast screening mammography[74,75].

Finally, our analysis highlights issues in the design of future studies investigating the biological basis of IBC. Specifically, there is a strong rationale for stratification of breast cancer cases based on measured breast density, specifically allowing for comparing subgroups to address confounding from mammographic density (*i.e.* study of ductal in low density breast, ductal in high density breasts, lobular in low density breast, lobular in high density breasts, each in IBC vs SDBC).

In conclusion, accepting the above caveats our findings indicate that while IBCs may not represent a distinct molecular subtype of breast cancer they are characterised by a more aggressive phenotype, in part likely to be a consequence of the timing of tumour initiation. Given that systematic review of national breast screening programs has found that IBCs are linked to worse patient survival[76], as well as being pertinent to patient management, our findings have relevance to informing screening programs with respect to defining screening intervals.

METHODS

Patients

The 100kGP project (release v14) provides WGS data on tumour-normal pairs from breast cancer patients recruited through 13 Genomic Medicine Centres across England. Informed written consent was obtained for all participants. We restricted our WGS analysis to samples with high-quality data from PCR-free, flash-frozen fresh tumour samples (**Supplementary Methods**). Demographic and clinical data were obtained from Public Health England's National Cancer Registration and Analysis Service (PHE-NCRAS), NHS Digital and the Genomic Medicine Centres. Tumour pathology information was obtained from histology reports. After QC (**Supplementary Table 1**), 833 of the 1,488 patients with a diagnosis of breast cancer had a documented NHS screening history. To minimise bias associated with screening indication, we confined our analysis to breast cancers diagnosed through the NHS breast screening programme of women aged between 47 and 70 years (**Table 1**). Considering hormone

receptor status, we confined our analysis to ductal histology since no patients with triple negative (TN) tumours had lobular breast cancer.

Calling of somatic variants

In addition to using variant calls from the 100kGP analysis pipeline we: (i) removed alignment bias introduced by ISAAC soft clipping of semi-aligned reads[11]; (ii) called tumour copy number using Battenberg[12]; (iii) called structural variants (SVs) from a consensus of Manta[13], LUMPY[14], and DELLY[15]; (iv) removed insertions and deletions within 10bp of a common germline indel (**Supplementary Methods**)[16].

Driver gene identification

To compare the frequency and clonality of driver gene mutations between IBCs and SDBCs we considered all driver genes identified across the 100kGP cohort of 1,488 invasive primary breast cancers. We identified cancer driver genes in these 1,488 cases using IntOGen[17], which combines seven computational methods to detect signals of positive mutational selection of missense mutations in coding regions of the genome (**Supplementary Methods**). Details of pre-processing of mutations, combining driver gene identification methodologies, post-processing and annotation of driver gene mutations are provided in **Supplementary Methods**.

Signature analysis

de novo extraction of single-base-substitution (SBS), doublet-base-substitution (DBS), insertion and deletion (ID) signatures, copy number (CN) signatures and structural variant (SV) signatures including decomposition to known COSMIC signatures[18] (v3.3), was performed using SigProfilerExtractor[19]. We used the extraction performed by *Everall et al.*[20] and complemented this analysis by screening for evidence of homologous recombination deficiency (dHR) in tumours using HRDetect[21]; using the advocated probability threshold of 0.7 to classify tumours as exhibiting dHR.

Estimation of clonality

The clonal state of driver mutations were estimated by MutationTimeR[22]. Sample odds ratios (ORs) were calculated for both early/late and clonal/subclonal driver mutations with associated *P*-values calculated using Fisher's exact test. The ratio of subclonal to clonal mutations was used as a proxy for intratumor heterogeneity.

Immune evasion

Neoantigens were identified using pVAC-Seq[23] by predicting the binding affinities of epitopes that arise as a result of non-synonymous mutations, based on the HLA-alleles typed by POLYSOLVER[24]. We investigated three possible immune escape mechanisms, specifically: (i) non-synonymous mutation in any of the three (HLA-A,-B,-C) HLA Class I genes; (ii) loss of heterozygosity (LOH) in any of the three HLA-I genes or (iii) any inactivating mutation in one of the 22 antigen-presenting genes (APGs) involved in the IFN-γ pathway, the PF-L1 receptor, the CD58 receptor, and epigenetic escape via *SETDB1* (*APLNR, B2M, CANX, CALR, CD274,*

CD58, CIITA, ERAP1, ERAP2, IRF2, IFNGR1, IFNGR2, JAK1, JAK2, NLRC5, PDIA3, RFX5, SETDB1, STAT1, TAPBP, TAP1, TAP2)[25,26]. We considered a tumour sample with positive immune escape status if it exhibited any one of (i)-(iii). HLA gene mutations were found using POLYSOLVER and LOH at HLA was predicted using LOHHLA[27]. We estimated numbers of tumour infiltrating lymphocytes based on the somatic copy number in conjunction with read depth of the T-cell receptor-α gene[28].

Germline variants and polygenic scores

We examined the germline of the patients for pathogenic variants in any of the 14 well-established hereditary breast ovarian cancer (HBOC) susceptibility genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CHEK2*, *MSH6*, *PALB2*, *PTEN*, *RAD51C*, *RAD51D*, *STK11* and *TP53*). Assignment of pathogenicity was based on a CADD[29] score >30 and ClinVar annotation[30]. Genomic positions of canonical gene transcripts were retrieved from the Ensembl database (EnsDb.Hsapiens.v86)[31] and were referenced to build GRCh38.

To generate polygenic scores (PGS) (**Supplementary Methods**) we used genome-wide association studies (GWAS) summary statistics estimated in European populations for breast cancer risk reported by Mavaddat *et al*[32]. For the well-established modifiable risk factors for breast cancer, we used results from the following resources: GSCAN consortium meta-analysis of smoking initiation (ever vs never status)[33] UK biobank (UKBB) meta-analysis of body mass index (BMI)[34], summary statistics relating to breast density reported by Chen *et al*.[35], and those relating to diabetes, such as fasting glucose and fasting insulin, were obtained from UKBB studies[36].

Statistical analysis

The relationship between categorical variables was assessed using either Chi-square or Fisher exact tests. Linear and logistic regression were used in analysis of continuous traits. The Kolmogorov–Smirnov test was performed to compare cumulative differences in PGS profiles. All statistical analysis we performed using R Version 4.2 and we considered a two-sided *P*-value <0.05 as statistically significant

DATA AVAILABILITY

The data supporting the findings of this study are available within the Genomics England Research Environment, a secure cloud workspace. An example for details on how to access data for this publication can be found at https://re-docs.genomicsengland.co.uk/pan_cancer_pub/. Additional processed aggregated data supporting the findings presented in this manuscript can be found in the Supplementary Tables. To access genomic and clinical data within this Research Environment, researchers must first apply to become a member of either the Genomics England Research Network (https://www.genomicsengland.co.uk/research/academic) or the Discovery Forum (industry partners https://www.genomicsengland.co.uk/research/research-environment). The process for joining the network is described at https://www.genomicsengland.co.uk/research/academic/join-gecip and consists of the following steps:

1. Your institution will need to sign a participation agreement available at https://files.genomicsengland.co.uk/documents/Genomics-England-GeCIP-Participation-Agreement-v2.0.pdf and email the signed version to gecip-help@genomicsengland.co.uk.

2. Once you have confirmed your institution is registered and have found a domain of interest, you can apply through the online form at https://www.genomicsengland.co.uk/research/academic/join-gecip. Once your Research Portal account is created you will be able to login and track your application.

3. Your application will be reviewed within 10 working days.

4. Your institution will validate your affiliation.

5. You will complete online Information Governance training and will be granted access to the Research Environment within 2 days of passing the online training.

Data that has been made available to registered users include: alignments in BAM or CRAM format, annotated variant calls in VCF format, signatures assignment, tumour mutation burden, sequencing quality metrics, summary of findings that is shared with Genomic Lab Hubs, secondary clinical data as described in this paper. Further details of the types of data available (for example, mortality, hospital episode statistics and treatment data) can be found at https://re-docs.genomicsengland.co.uk/data_overview/. Germline variants can be explored in Interactive Variant Analysis Browser (see description at https://re-docs.genomicsengland.co.uk/iva_variant/). Cancer patients cohort and longitudinal clinical information on treatment and mortality can be explored with Participant Explorer (see description at https://re-docs.genomicsengland.co.uk/pxa/).

CODE AVAILABILITY

Code used in the generation of results and figures reported in this study can be found within the Genomics England Research Environment in the following directory: /re_gecip/shared_allGeCIPs/cmills/publications/breast_interval

COMPETING INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

A.S., D.C. and R.S.H. designed the study. A.S., A.J.C. and D.C. performed sample curation, C.M., A.S., A.E., D.C., S.L., B.K., A.J.C. and R.B. performed bioinformatic and statistical analysis. C.M. A.S., A.E. and R.S.H. drafted the manuscript; all authors reviewed, read, and approved the final manuscript. C.M., A.S. and A.E. are considered co-first authors.

REFERENCES

1. Pisano, E. D. *et al.* Diagnostic performance of digital versus film mammography for breast-cancer screening. *N. Engl. J. Med.* **353**, 1773–1783 (2005).

2. Shapiro, S., Strax, P. & Venet, L. Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA* **215**, 1777–1785 (1971).

3. Tabár, L. *et al.* Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* **1**, 829–832 (1985).

4. *Screening for Breast Cancer: Systematic Evidence Review Update for the U.S. Preventive Services Task Force*. (2009).

5. Bennett, R. L., Sellars, S. J. & Moss, S. M. Interval cancers in the NHS breast cancer screening programme in England, Wales and Northern Ireland. *Br. J. Cancer* **104**, 571–577 (2011).

6. Ma, L. *et al.* Case-control study of factors associated with failure to detect breast cancer by mammography. *J. Natl. Cancer Inst.* **84**, 781–785 (1992).

7. Timmermans, L. *et al.* Tumour characteristics of screen-detected and interval cancers in the Flemish Breast Cancer Screening Programme: A mammographic breast density study. *Maturitas* **158**, 55–60 (2022).

8. Houssami, N. & Hunter, K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer* **3**, 12 (2017).

9. Cheasley, D. *et al.* Molecular comparison of interval and screen-detected breast cancers. *J. Pathol.* **248**, 243–252 (2019).

10. Li, J. *et al.* Molecular Differences between Screen-Detected and Interval Breast Cancers Are Largely Explained by PAM50 Subtypes. *Clin. Cancer Res.* **23**, 2584–2592 (2017).

11. Cornish, A. J. *et al.* Reference bias in the Illumina Isaac aligner. *Bioinformatics* vol. 36 4671–4672 (2020).

12. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).

13. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

14. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).

15. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

16. Kinnersley, B. *et al.* Cancer driver genes and opportunities for precision oncology revealed by whole genome sequencing 10,478 cancers. *medRxiv* 2023.05.24.23289454 (2023) doi:10.1101/2023.05.24.23289454.

17. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).

18. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2018).

19. Ashiqul Islam, S. M. *et al. Uncovering Novel Mutational Signatures by de Novo Extraction with SigProfilerExtractor*. (2022).

20. Everall, A. *et al.* Comprehensive repertoire of the chromosomal alteration and mutational signatures across 16 cancer types from 10,983 cancer patients. *bioRxiv* (2023) doi:10.1101/2023.06.07.23290970.

21. Davies, H. *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).

22. Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).

23. Hundal, J. *et al.* pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **8**, 11 (2016).

24. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).

25. Martínez-Jiménez, F. *et al.* Genetic immune escape landscape in primary and metastatic cancer.

Nat. Genet. **55**, 820–831 (2023).

26. Kelly, A. & Trowsdale, J. Genetics of antigen processing and presentation. *Immunogenetics* **71**, 161–170 (2019).

27. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171**, 1259–1271.e11 (2017).

28. Bentham, R. *et al.* Using DNA sequencing data to quantify T cell fraction and therapy response. *Nature* **597**, 555–560 (2021).

29. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).

30. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–5 (2014).

31. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).

32. Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Inst.* **107**, (2015).

33. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).

34. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

35. Chen, H. *et al.* Genome-wide and transcriptome-wide association studies of mammographic density phenotypes reveal novel loci. *Breast Cancer Res.* **24**, 27 (2022).

36. Mansour Aly, D. *et al.* Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat. Genet.* **53**, 1534–1542 (2021).

37. Kinnersley, B. *et al.* Opportunities for precision oncology revealed by whole genome sequencing 10,478 cancers - the UK 100,000 Genomes Project. *bioRxiv* (2023) doi:10.1101/2023.05.24.23289454.

38. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).

39. Weber, R. J. P. *et al.* Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. *Breast Cancer Res. Treat.* **158**, 471–483 (2016).

40. Meshkat, B. *et al.* A comparison of clinical-pathological characteristics between symptomatic and interval breast cancer. *Breast* **24**, 278–282 (2015).

41. Holm, J. *et al.* Risk factors and tumor characteristics of interval cancers by mammographic density. *J. Clin. Oncol.* **33**, 1030–1037 (2015).

42. Domingo, L. *et al.* Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. *Breast Cancer Res.* **16**, R3 (2014).

43. Renart-Vicens, G. *et al.* Evaluation of the interval cancer rate and its determinants on the Girona Health Region's early breast cancer detection program. *BMC Cancer* **14**, 558 (2014).

44. Boyd, N. F. *et al.* Mammographic features associated with interval breast cancers in screening programs. *Breast Cancer Res.* **16**, 417 (2014).

45. Caldarella, A. *et al.* Biological characteristics of interval cancers: a role for biomarkers in the breast cancer screening. *J. Cancer Res. Clin. Oncol.* **139**, 181–185 (2013).

46. Payne, J. I., Caines, J. S., Gallant, J. & Foley, T. J. A review of interval breast cancers diagnosed among participants of the Nova Scotia Breast Screening Program. *Radiology* **266**, 96–103 (2013).

47. Kalager, M., Tamimi, R. M., Bretthauer, M. & Adami, H.-O. Prognosis in women with interval breast cancer: population based observational cohort study. *BMJ* **345**, e7536 (2012).

48. Caumo, F. *et al.* Interval cancers in breast cancer screening: comparison of stage and biological characteristics with screen-detected cancers or incident cancers in the absence of screening. *Tumori* **96**, 198–201 (2010).

49. Hofvind, S., Geller, B. & Skaane, P. Mammographic features and histopathological findings of interval breast cancers. *Acta radiol.* **49**, 975–981 (2008).

50. Baré, M. *et al.* Interval breast cancers in a community screening programme: frequency, radiological classification and prognostic factors. *Eur. J. Cancer Prev.* **17**, 414–421 (2008).

51. Evans, A. J. *et al.* Radiological and pathological findings of interval cancers in a multi-centre, randomized, controlled trial of mammographic screening in women from age 40-41 years. *Clin. Radiol.* **62**, 348–352 (2007).

52. Porter, G. J. R. *et al.* Interval breast cancers: prognostic features and survival by subtype and time since screening. *J. Med. Screen.* **13**, 115–122 (2006).

53. Collett, K. *et al.* A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1108–1112 (2005).

54. Eriksson, L. *et al.* Mammographic density and survival in interval breast cancers. *Breast Cancer Res.* **15**, R48 (2013).

55. Gilliland, F. D. *et al.* Biologic characteristics of interval and screen-detected breast cancers. *J. Natl. Cancer Inst.* **92**, 743–749 (2000).

56. Vitak, B. *et al.* Interval cancers and cancers in non-attenders in the Ostergötland Mammographic Screening Programme. Duration between screening and diagnosis, S-phase fraction and distant recurrence. *Eur. J. Cancer* **33**, 1453–1460 (1997).

57. Rodriguez, J. *et al.* Investigation of Genetic Alterations Associated With Interval Breast Cancer. *JAMA Oncol* (2024) doi:10.1001/jamaoncol.2023.6287.

58. Brekelmans, C. T. *et al.* Effectiveness of breast cancer surveillance in BRCA1/2 gene mutation carriers and women with high familial risk. *J. Clin. Oncol.* **19**, 924–930 (2001).

59. Tilanus-Linthorst, M. M. A. *et al.* BRCA1 mutation and young age predict fast breast cancer growth in the Dutch, United Kingdom, and Canadian magnetic resonance imaging screening trials. *Clin. Cancer Res.* **13**, 7357–7362 (2007).

60. Rijnsburger, A. J. *et al.* BRCA1-associated breast cancers present differently from BRCA2-associated and familial cases: long-term follow-up of the Dutch MRISC Screening Study. *J. Clin.*

Oncol. **28**, 5265–5273 (2010).

61. Chéreau, E. *et al.* Characteristics, treatment, and outcome of breast cancers diagnosed in BRCA1 and BRCA2 gene mutation carriers in intensive screening programs including magnetic resonance imaging. *Clin. Breast Cancer* **10**, 113–118 (2010).

62. Scheuer, L. *et al.* Outcome of Preventive Surgery and Screening for Breast and Ovarian Cancer in BRCA Mutation Carriers. *J. Clin. Oncol.* (2016) doi:10.1200/JCO.2002.20.5.1260.

63. Shah, P. *et al.* Prospective study of breast MRI in BRCA1 and BRCA2 mutation carriers: effect of mutation status on cancer incidence. *Breast Cancer Res. Treat.* **118**, 539–546 (2009).

64. Grassmann, F. *et al.* Interval breast cancer is associated with other types of tumors. *Nat. Commun.* **10**, 1–9 (2019).

65. Cheasley, D. *et al.* The TP53 mutation rate differs in breast cancers that arise in women with high or low mammographic density. *NPJ Breast Cancer* **6**, 34 (2020).

66. Reed, A. E. M., Kutasovic, J. R., Lakhani, S. R. & Simpson, P. T. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. *Breast Cancer Res.* **17**, 1–11 (2015).

67. Sieh, W. *et al.* Identification of 31 loci for mammographic density phenotypes and their associations with breast cancer risk. *Nat. Commun.* **11**, 5116 (2020).

68. Goldhirsch, A. *et al.* Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann. Oncol.* **22**, 1736–1747 (2011).

69. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).

70. Kaelin, W. G., Jr. The concept of synthetic lethality in the context of anticancer therapy. *Nat. Rev. Cancer* **5**, 689–698 (2005).

71. Kamb, A. Mutation load, functional overlap, and synthetic lethality in the evolution and treatment of cancer. *J. Theor. Biol.* **223**, 205–213 (2003).

72. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy.

Nature **434**, 917–921 (2005).

73. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).

74. Lv, R. *et al.* Noninvasive detection of chromosomal instability in plasma circulating cell-free DNA for early pan-cancer diagnosis using low-pass whole-genome sequencing. *J. Clin. Oncol.* **39**, e22509–e22509 (2021).

75. Lin, Z. *et al.* Abstract 5162: TuFEst: a sensitive and cost-effective pan-cancer detection approach with accurate tumor fraction estimation. *Cancer Res.* **82**, 5162–5162 (2022).

76. Canelo-Aybar, C. *et al.* Benefits and harms of annual, biennial, or triennial breast cancer mammography screening for women at average risk of breast cancer: a systematic review for the European Commission Initiative on Breast Cancer (ECIBC). *Br. J. Cancer* **126**, 673–688 (2021).

FIGURE CAPTIONS

Figure 1. Study overview. WGS, whole genome sequencing. Created with BioRender.

Figure 2. Mutational burden in IBC and SDBC. The mutational burden distribution for IBCs and SDBCs in selected cohorts. The median of the distribution is indicated by the horizontal line and *P*-values calculated using a Wilcoxon rank sum test.

Figure 3. Frequency and clonality of driver gene mutations in IBC and SDBC. The mutational frequency of driver genes for: (**a**) all IBCs and SDBCs; (**b**) ductal IBCs and SDBCs (**c**) lobular IBCs and SDBCs. The colour scale corresponds to the total fraction of detected mutations that are clonal, with a fraction of 1 indicating that every detected mutation is clonal and a fraction of 0 indicating that every detected mutation is subclonal. The clonal fraction is defined as $N_{clonal}/(N_{clonal}+N_{subclonal})$, where N is the number of detected mutations in the gene across the cohort. ** significant difference in mutational frequency imposing a Bonferroni adjusted *P*-value of 7.46 x 10^{-4}, * a significant difference at an unadjusted *P*-value of 0.05. § a significant difference in clonal fraction at an unadjusted *P*-value of 0.05.

Figure 4. Copy number profile of IBC and SDBC. (a) IBC, (**b**) SDBC. Loss of heterozygosity is indicated in light blue, while other copy number losses are displayed in dark blue. Copy number gains are shown in red while large amplifications are indicated in orange. The regions with focal alterations, as defined by GISTIC2, are annotated with grey bars.

Figure 5. Proportion of the genome altered in IBC and SDBC. The distribution of the fraction of genome altered for IBCs and SDBCs. A region of the genome is considered to be altered if there is any deviation from a single copy of the major and minor allele. The median of the distribution is indicated by the horizontal line and *P*-values are calculated using a Wilcoxon rank sum test.