Mutational landscape of pure ductal carcinoma in situ and associations with disease prognosis and response to radiotherapy

Noor Rizvi[1], Eliseos J. Mucaki[1,2], Emily L. Salmini[1], Monica Zhang[2], Sabina Trebinjac[3], Ezra Hahn[3], Lawrence Paszat[3], Sharon Nofech-Mozes[4], Michael T. Hallett[1,5,6], Eileen Rakovitch[3] , Vanessa Dumeaux[1,2,5,6]

[1] Department of Biochemistry, Western University; [2] Department of Anatomy and Cell Biology, Western University; [3] Department of Radiation Oncology, Sunnybrook Health Sciences Centre, University of Toronto; [4] Department of Pathology, Sunnybrook Health Sciences Centre, University of Toronto; [5] Department of Oncology, Lawson Research Institute, London Health Sciences Centre; [6] Centre for Translational Cancer Research, Western University

Abstract

Background

Managing Ductal Carcinoma in Situ (DCIS) remains challenging due to the lack of reliable biomarkers to predict radiotherapy (RT) response, leading to both overtreatment of indolent disease and undertreatment of aggressive cases.

Results

Through whole-exome sequencing of 147 DCIS cases, we characterized the genomic landscape of pure DCIS and identified genetic alterations associated with the risk of recurrence, either in-situ or invasive. DCIS lesions harbored frequent mutations in established cancer drivers (*PIK3CA*, *TP53*) and genes regulating tissue architecture, which likely enhanced pre-invasive cell fitness but lacked prognostic value. A subset of younger patients exhibited distinct mutational processes characterized by high mutational burden, though these were not linked to recurrence risk. Across the cohort, five mostly mutually exclusive genes (*SH2B2*, *PDZD8*, *MYO7A*, *MUCL3*, *DNASE2B*), involved in cell adhesion, membrane organization, and DNA degradation, were significantly associated with early recurrence (within 5 years). In RT-treated patients, we identified 27 additional mutated genes uniquely associated with recurrence, along with *SH2B2* and *MUCL3*, which were also identified in the whole cohort. Most of these genes are involved in cytoskeletal regulation, cell adhesion, and cell-environment interactions. Mutations in two metabolic regulators (*MGAM2* and *AADACL3*), as well as *REV1*, a key player in DNA damage tolerance, may impair cellular responses to RT-induced stress or genomic stability. Copy number analyses revealed that pure DCIS exhibits molecular subtype-specific patterns characteristic of invasive disease, with novel recurrence-associated

alterations, including three non-adjacent gains and 12 losses in regions harboring oncogenes, tumor suppressor genes, and genes regulating structural integrity, cell-cell adhesion and interactions.

Conclusions

Our findings identify key genetic alterations associated with local recurrence and RT resistance. While *TP53*, *PIK3CA*, and recurrent copy number alterations represent early events in tumorigenesis, they lack prognostic value in pure DCIS, underscoring the need for alternative biomarkers. By uncovering mutations associated with cytoskeletal integrity, cell adhesion, and cell-cell interactions, our study provides insights into DCIS recurrence and RT response, with potential to guide personalized therapeutic strategies.

Introduction

Ductal Carcinoma in Situ (DCIS) is a non-invasive, non-obligate precursor of invasive breast cancer characterized by clonal proliferation of neoplastic cells confined within the breast ducts [1]. DCIS is primarily detected through screening and diagnostic mammograms due to its characteristic presentation with microcalcifications. The widespread implementation of mammographic screening programs has led to a substantial increase in DCIS detection, with incidence rates rising from four to eleven cases per 100,000 women between 1993 and 2007 [2,3]. This trend is expected to continue as screening programs expand to include younger women [4].

There are currently no definitive markers to predict which cases will progress to invasive, life-threatening lesions [5,6]. Consequently, treatment, usually involving breast-conserving surgery (BCS) followed by breast radiation therapy (RT), is recommended for all women diagnosed with DCIS [5]. However, many DCIS would not transit to invasive life-threatening disease even if left untreated [7]. The absence of known markers predictive of a patient benefit to RT results in both over-treatment for indolent lesions and under-treatment for some aggressive DCIS likely to evolve to an invasive state. Markers that identify such cases could enable tailored treatments, such as full mastectomy or adjuvant systemic therapies for patients at risk of in situ or invasive local recurrence (LR) despite receiving RT [8], or more frequent watch-and-wait strategies for those with indolent lesions.

Cancer emerges from the accumulation of genetic aberrations in neoplastic cells and genomic instability [9]. Mutations can disrupt normal cellular processes, leading to uncontrolled cell growth, impaired DNA repair, and other aberrant cellular behaviours that may contribute to cancer development [9]. Previous studies have revealed important insights into DCIS progression by highlighting specific mutations and chromosomal alterations that may influence the progression towards invasive disease [10–17]. Many of these studies however, focus on synchronous DCIS – concurrent presentation of DCIS and invasive ductal carcinoma (IDC). These studies explore the overlap between DCIS and IDC, and recognize the potential differences between them. However these findings might be reflective of a timepoint beyond the evolutionary bottleneck, and the full repertoire of mediators of the transition from DCIS to IDC or associated with a future risk of LR cannot be established from these studies [10].

TP53 and *PIK3CA* [11–17] are among the most frequently observed mutations in these studies [11–17] and are often characterized as key drivers of tumorigenesis, promoting growth and spread of cancer cells. However, this may not necessarily be the case, as mutant field clonalization could equip pre-malignant cells with additional (epi-) genetic mutations that confer fitness advantages, allowing them to expand into ductal regions without directly driving invasive behavior [10]. This suggests that driver mutations, alongside frequently mutated genes, may serve as these early fitness enhancers rather than drivers of progression. Several copy number aberrations (CNAs) have also been identified by these studies [10–15,17–20] including frequent gains at 1q, 8q, 11q, and 17q, and losses at 16q, however the specific associations of these CNAs to DCIS prognosis remain unclear.

To address these critical gaps, we conducted comprehensive exome sequencing analysis of 147 pure DCIS cases, including patients treated with and without RT, to investigate markers of local recurrence within 10 years of diagnosis. Our study reveals a distinct mutational landscape in pure DCIS and identifies novel genomic alterations associated with tumor grade, molecular subtypes, and patient age. Most importantly, we discovered specific variants and CNAs predictive of local recurrence risk, including genetic markers associated with RT response. These findings provide insights into the genomic determinants of DCIS prognosis and treatment response, establishing a foundation for improved risk stratification and personalized treatment strategies for DCIS patients.

Results

A unique cohort of pure DCIS patients

We assembled the largest cohort of pure DCIS patients treated with BCS, with or without subsequent RT, incorporating comprehensive genomic profiling through whole-exome DNA analysis of primary DCIS tumors and matched normal tissues (n = 147; **Table 1**). The study design was balanced to include at least a third of the patients who experienced an ipsilateral invasive or in-situ local recurrence within a 10 year follow-up period. The median recurrence time in these patients was 4.2 years for invasive disease and 2.1 years for in-situ disease. Most tumors were of intermediate to high grade, spanning all five molecular subtypes, with no multifocality and negative margins (**Table 1**).

		Tumor-Normal pairs	
		Without Local Recurrence (n = 96)	**With Local Recurrence (n = 51)**
Time to Recurrence in years Median (Range)	DCIS	NA	2.1 (0.5 – 9.7)
	Invasive	NA	4.2 (0.6 – 9.7)
Radiotherapy (RT)	Yes	51	22
	No	45	29
Clear Margins	Positive	5	5
	Negative	79	40
	Undetermined	12	6
Tumor size (mm), Median (Range)		15 (4 - 90)	18 (2 - 76)
Nuclear Grade	Low	2	2
	Moderate	61	26
	High	33	23
Multifocality	Present	19	17
	Absent	48	24
	Undetermined	29	10
Age	<50 Years Old	37	15
	50-60 Years Old	34	14
	>60 Years Old	25	22
PAM50 Subtype (Pearson Correction)	Basal-like	12	9
	Her2-like	9	9
	LumA	27	5
	LumB	9	12
	Normal-Like	20	5

Table 1. Patient and tumor clinical attributes

Mutational landscape of DCIS reveals distinct processes driving high mutational burden in early-onset cases

The mutational load of pure DCIS lesions varied considerably, ranging from 3 to 3,482 variants per sample, with a median of 75 non-synonymous variants. The majority (74.2%) were missense mutations, constituting approximately 52K unique variants identified in ~14K genes.

As expected, most mutations are of C>T type (barplot in **Fig. 1**), a common mutational pattern attributed to the spontaneous deamination of 5-methylcytosine, a process frequently observed in many cancer genomes [21,22]. Spontaneous deamination can be exacerbated in FFPE samples due to DNA damage [23]. FFPE-related mutational artefacts are known to resemble certain COSMIC signatures [24], such as SBS30 and SBS1 [25]. During library preparation, formalin-induced DNA lesions are chemically repaired with unrepaired profiles resembling SBS30 and repaired profiles resembling SBS1 [25]. While SBS30 was not identified in our dataset, SBS1 was detected in a large number of samples (**Fig. 1)**. SBS1, commonly found in tumor genomes, is difficult to distinguish from repaired FFPE-related artifacts due to their high similarity [22,25]. However, previous research has demonstrated that mutational profiles from repaired FFPE samples closely match true tumor mutational profiles [25], and we expect that our careful processing of mutation calling-pipeline minimizes the impact of these artefacts [23].

Interestingly, a subset of DCIS samples exhibit a significantly higher tumor mutational burden (TMB), characterized by an increased in frequency of C>G, T>C, and C>A mutations, and enrichment of specific SBS signatures including SBS26, SBS7b, and SBS5 (**Fig. 1**, 12 samples on the right). The causes underlying these mutational signatures remain incompletely

understood. However, SBS26 has been linked to impaired mismatch repair and microsatellite instability. These cases were not more likely to have a recurrence, of a higher grade or of a specific molecular subtype but were predominantly found in younger patients (8 out of 12 under 50 years; chisq.test p-value < 0.05).

These findings highlight distinct mutational processes in some early-onset DCIS cases, though these molecular features were not associated with prognosis.

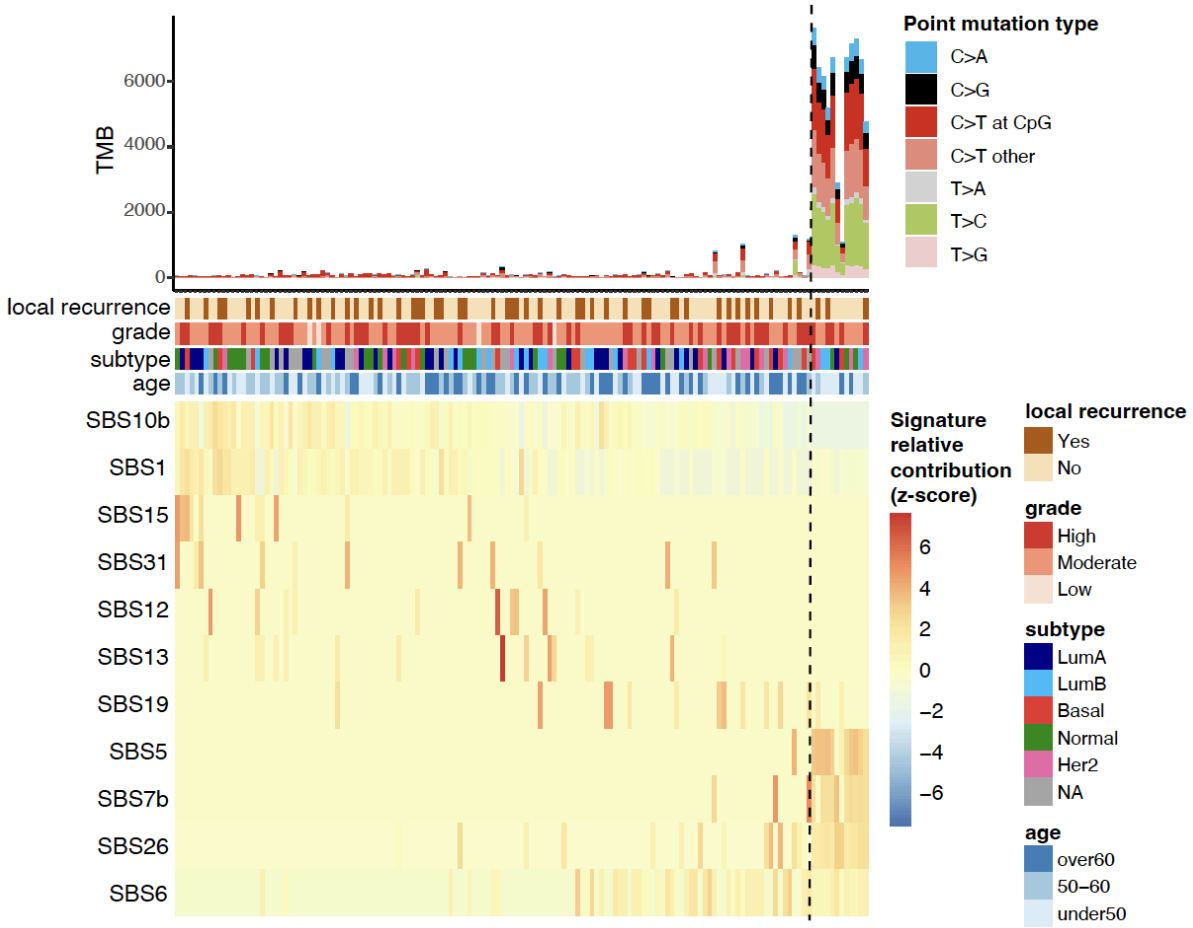

Fig. 1: Mutational Signatures in pure DCIS samples.

Top panel shows tumor mutational burden (TMB) with distribution of point mutation types. Clinical annotations display 10-year local recurrence outcome, tumor grade, molecular subtype,

and patient age. The bottom heatmap depicts the relative contribution across samples of single base substitution (SBS) signatures detected in at least 10 samples (z-score normalized).

Pure DCIS is associated with high frequency of mutations in genes involved cell adhesion, polarity, tissue structure and function

Analysis of pure DCIS revealed distinct patterns of recurrent mutations across multiple genes (**Fig. 2**). *PIK3CA* was the most frequently mutated gene (15% of cases), followed by *FSIP2* and *KIR3DL3* (14%). We also identified functional gene groups among the most frequently mutated genes including several motor genes converting chemical energy to mechanical force (*DNAH12, DNHD1*, and *MYOB15;* 12% each), collagen genes (*COL18A1* and *COL4A3*; 12% and 10%, respectively), and mucin genes (*MUC3A*, *MUC4*, *MUC22*, and *MUC5AC)* also showed alterations (10-12% of cases each). While mucin genes are typically large and can accumulate mutations by chance, these specific mucin genes were not identified as FLAGS (FrequentLy mutAted GeneS) and therefore not excluded in our analysis (See Methods). Given our stringent rules for mutation calling, these findings suggest that mutations in mucin and other genes controlling epithelial-components including cellular morphology, epithelial function and adhesion are central to the physiopathology of pure DCIS.

We also identified a few mutated genes enriched in distinct clinico-pathological groups including patients with early-onset DCIS (*FILP1L*, *CFAP61*, *FREM1*) or later-onset (*ERBB4*), high-grade lesions (*TP53*) and Her2-enriched subtype (*ASH1L, NAGPA, DMD*) (**Figure 2B**).

Collectively, these findings highlight that pure DCIS harbors frequent mutations in genes governing tissue architecture and cell-cell interactions, suggesting these alterations may be fundamental to DCIS development.

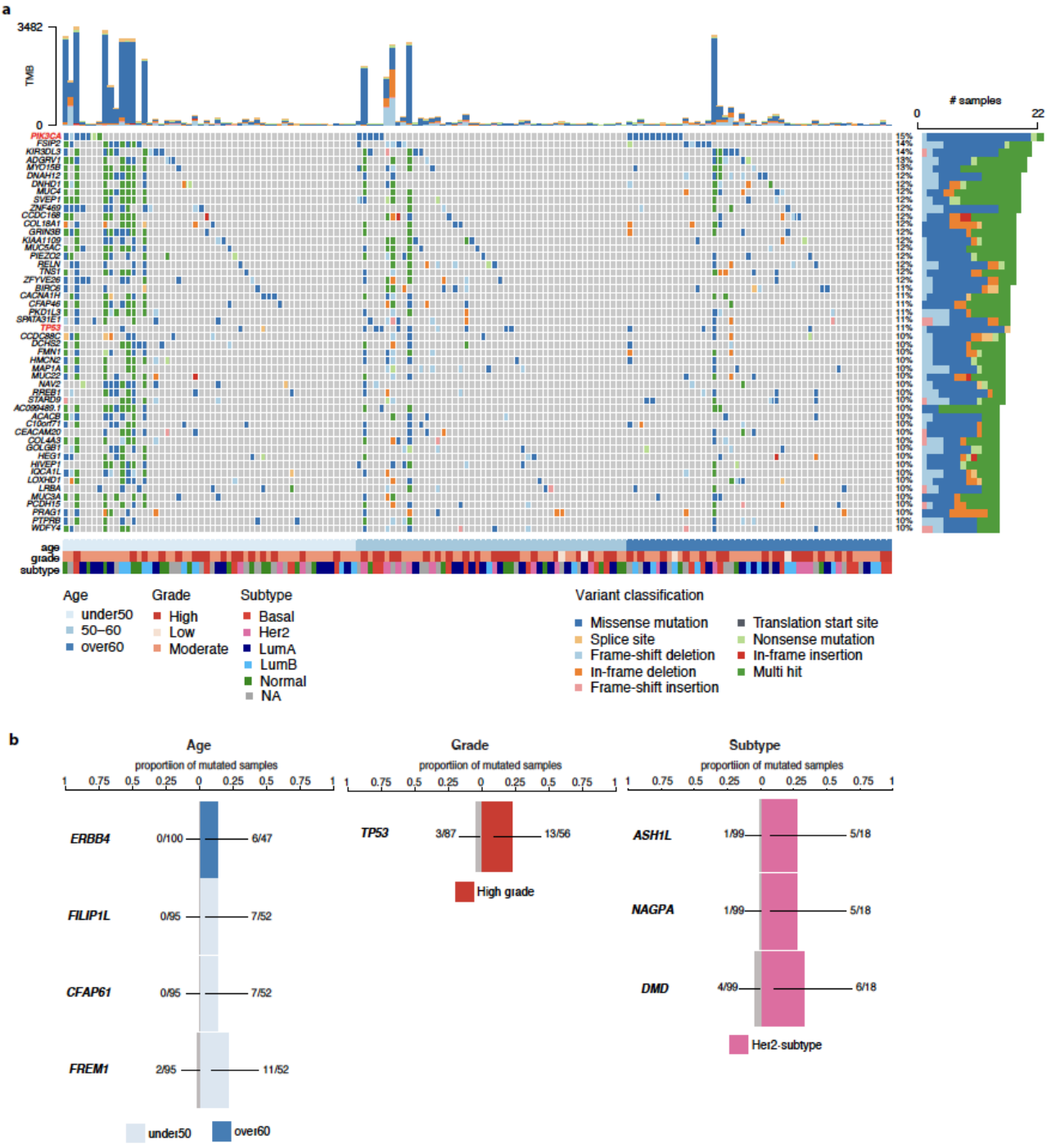

Fig. 2: The most frequently mutated genes and association with clinico-pathological variables in pure DCIS patients.

a The top 50 most frequent non-synonymous small variants identified in pure DCIS lesions. Samples are in columns and variants are color-coded based on their classification. The TMB for each lesion is displayed at the top of the heatmap. Samples are organized by age at diagnosis, with additional clinico-pathological features—grade and subtype—depicted at the bottom. **b** Mutated genes significantly associated with specific clinico-pathological variables (False discovery rate [FDR] < 0.001). The barplots show the proportion of mutated samples within specific categories: age group (left), high grade (middle), and Her2-subtype (right). Proportions for each category are compared to the proportions of mutated samples in the other respective groups (shown in grey).

Established cancer driver genes are present in pure DCIS but lack prognostic capacity

Mutational analysis identified two significant driver genes in pure DCIS: *PIK3CA* and *TP53* (FDR < 0.05). Both genes were also among the most frequently mutated genes (**Fig. 2a**, red-labelled genes). Out of 34 samples carrying a mutation in at least one of these driver genes, 15 samples harbored at least one potentially actionable alteration as indicated by OncoKB (**Supplementary Fig. 1**).

The mutation spectrum in *PIK3CA* was dominated by the activating H1047R hotspot mutation (~50%) in the kinase domain of exon 21, known to enhance PI3K-mitigated pathway signaling (**Supplementary Fig. 1a**) [26,27]. Similarly, *TP53* mutations clustered in known hotspots within

the DNA-binding domain, which may result in the loss of tumor suppression by affecting its ability to bind to DNA (**Supplementary Fig. 1b**).

Despite their established roles in cancer progression, none of these driver mutations correlated with increased 10-year local recurrence risk, suggesting that additional factors influence DCIS prognosis.

Identification of mutations associated with increased local recurrence risk independent of treatment

To identify potential prognostic markers, we analyzed mutated genes associated with the risk of invasive or in-situ local recurrence in the ipsilateral breast occurring between 6 months and 10 years after diagnosis. Our survival analysis identified five biomarkers significantly associated with increased 10-year local recurrence risk (log-rank test p-value < 0.01; **Fig. 3a-b**). These mutations, occurring in 4-7% of cases, were largely mutually exclusive and occurred across molecular subtypes and grades.

The presence of mutations in at least one of these genes was significantly associated with increased recurrence risk (**Fig. 3c,** log-rank p-value p<0.0001), with *MYO7A* and *PDZD8* showing the strongest associations (Firth's penalized likelihood Cox regression HR>4.4, **Fig. 3b**). The genes represent diverse cellular functions, clustering in three major functional groups: cell adhesion (*MYO7A, MUCL3*), organization and function of the cellular membrane (*SH2B2, PDZD8, MYO7A*), and DNA degradation (*DNASE2B*) (**Fig. 3d**). These findings underscore how

cytoskeletal reorganization, changes in cell structure, and compromised cell adhesion contribute to increased risk of recurrence within 5 years after a DCIS diagnosis.

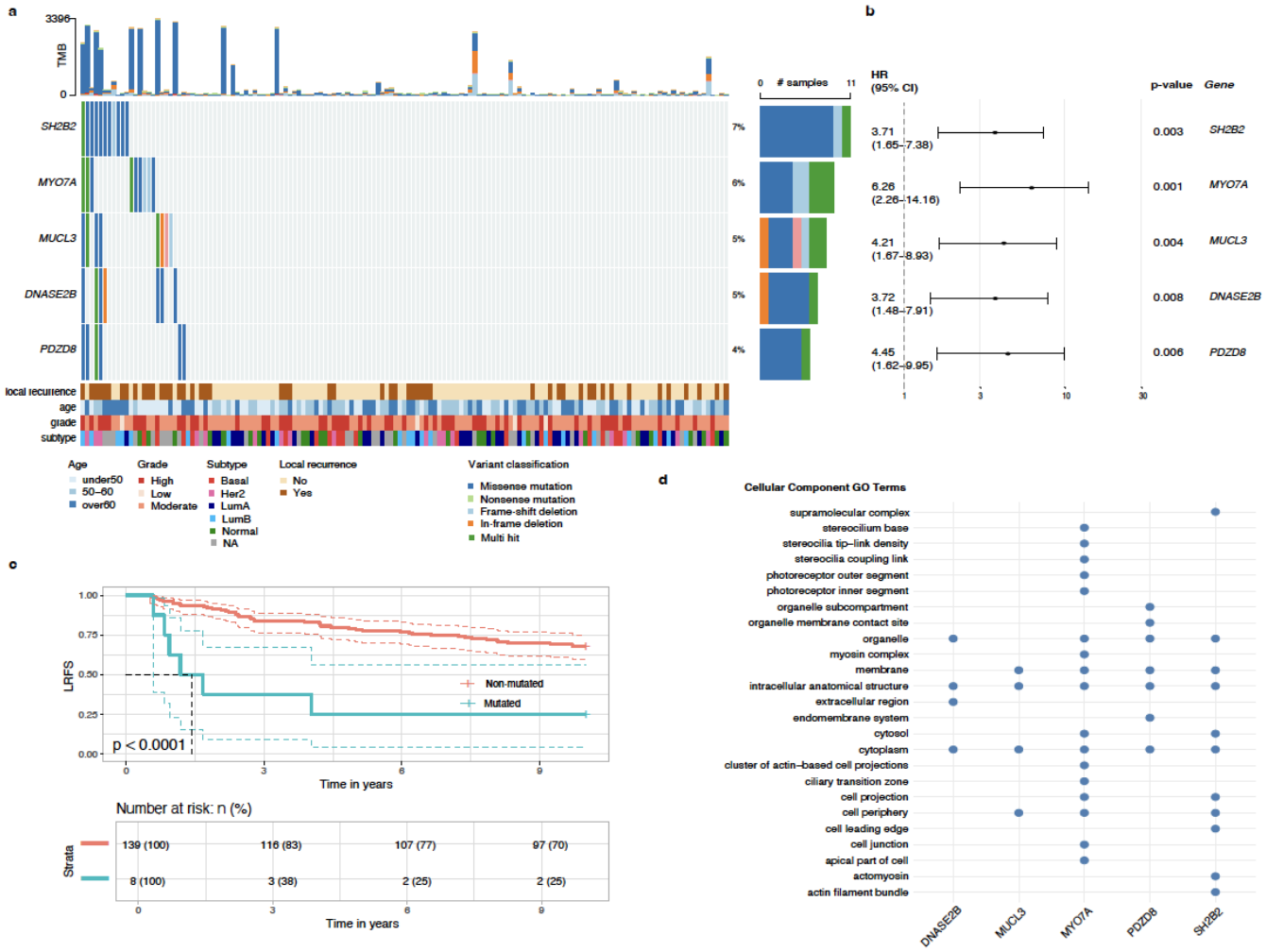

Fig. 3: Gene variants associated with an increased 10-year local recurrence risk in DCIS. **a** Samples are in columns and variants are color-coded based on their classification. Local recurrence status and clinico-pathological characteristics of each lesion are depicted at the bottom. Local recurrence is defined as any recurrence, in situ or invasive, in the ipsilateral breast occurring between 6 months and 10 years after diagnosis. Right-hand side reports the proportion of each variant classification type. **b** Hazard ratio and confidence intervals for each

significant mutated gene associated with increased 10-year local recurrence risk (Firth's penalized likelihood Cox regression). **c** Kaplan-Meier analysis of local recurrence-free survival (LRFS) comparing patients with mutations in at least one of the 5 genes (red) versus those without mutations (blue). **d** Cellular component Gene Ontology (GO) terms annotations for each significant gene.

Mutations in genes governing cytoskeletal organization and membrane dynamics associated with radiotherapy resistance

To identify predictive biomarkers for RT response, we analyzed mutations associated with 10-year local recurrence in a cohort restricted to patients who received RT. This analysis revealed 29 genes significantly associated with an increased risk of recurrence (log-rank test $p < 0.05$; **Fig. 4a**). Notably, these mutations often co-occurred, with at least one mutated gene present in approximately 27% of lesions (20/73 patients treated with RT).

In RT-treated patients, lesions harboring mutations in at least one of these genes exhibited a markedly increased risk of local recurrence (Firth's penalized likelihood Cox regression HR = 4.8, 95% CI: 2.1–11.2, $p = 0.0002$), with most mutation-positive recurrences occurring within 5 years post-RT (**Fig. 4b**). Notably, while mutations in *SH2B2* and *MUCL3* were significantly associated with prognosis in both the overall cohort and the RT-treated subgroup, an analysis restricted to lesions harboring mutations in the remaining 27 genes revealed that these mutations were predictive of prognosis only in RT-treated patients ($p < 0.0001$) and not in those who did not receive RT ($p > 0.01$). We did not identify any significant mutations that were exclusively associated with prognosis in patients who did not receive RT.

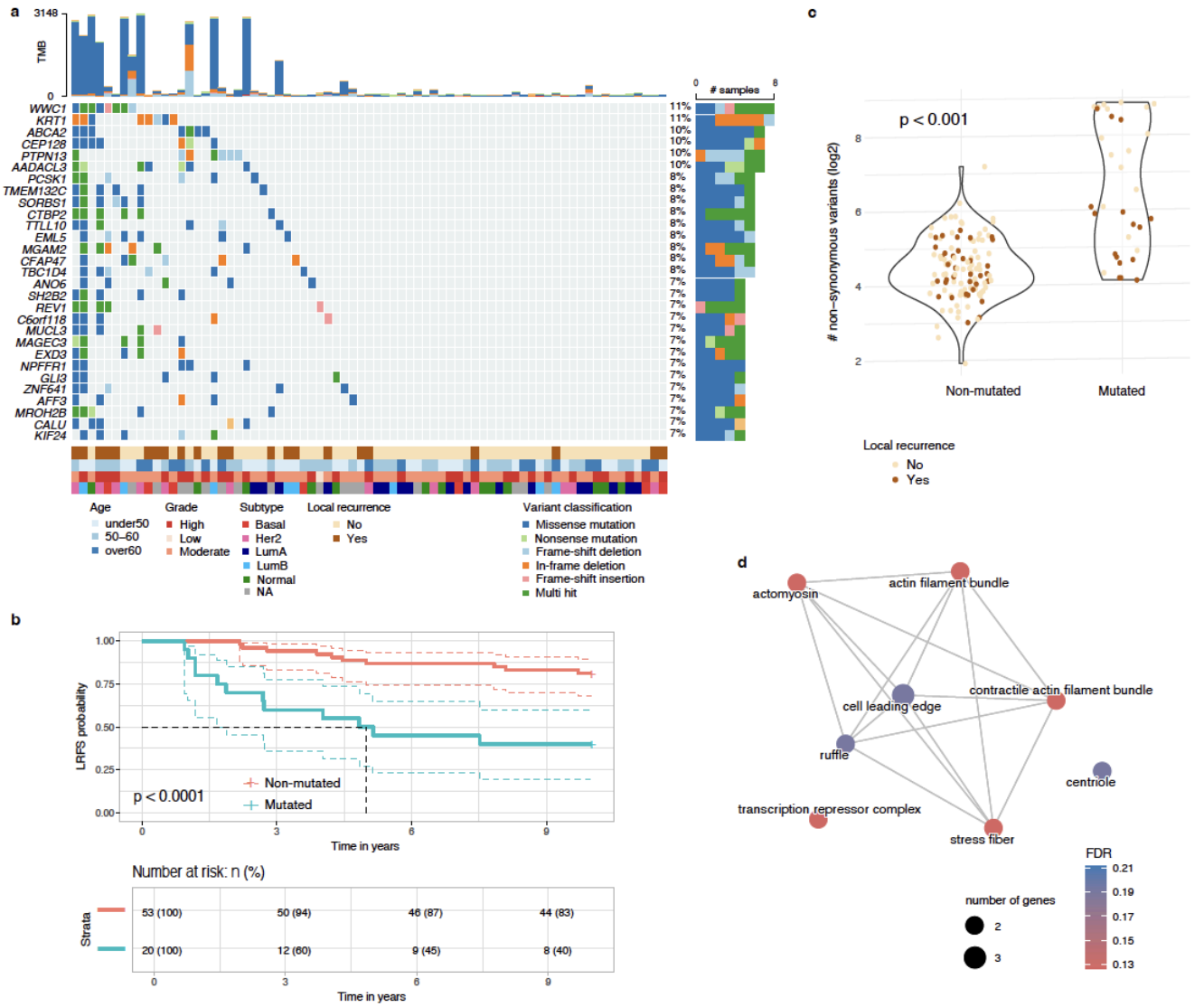

Fig. 4: Mutations associated with radiotherapy resistance in DCIS.

a An oncoplot showing the distribution of mutations in 29 genes significantly associated with local recurrence in RT-treated patients. Color-coding indicates mutation types; clinical annotations show age, grade, molecular subtype, and recurrence status. **b** Kaplan-Meier analysis of LRFS comparing patients with mutations in at least one of the 29 genes (red) versus those without mutations (blue). **c** Distribution of TMB, shown as number of

non-synonymous variants, log2 scale) for lesions with at least one mutation in genes associated with LRFS colored by recurrence status at 10 years. **d** GO cellular component enrichment network analysis of recurrence-associated genes in RT-treated DCIS. Network visualization shows enriched GO terms (FDR < 0.2) from genes linked to local recurrence following RT. Nodes represent individual GO terms with size proportional to gene count (>2), and edges indicate significant semantic similarity between terms. Node color intensity corresponds to enrichment significance.

Given the frequent co-occurrence of these mutations, we assessed whether overall TMB might explain the increased recurrence risk. Although several high-TMB lesions were observed among patients with these mutations (**Fig. 4a,c**), TMB itself was not significantly associated with local recurrence risk. Indeed, many lesions with high TMB remained recurrence-free over 10 years, suggesting that specific mutations—not overall mutation load—drive LR after RT.

We further examined potential confounding by clinico-pathological variables. After adjusting for age, the predictive value of these mutations remained significant (Firth's penalized likelihood Cox regression HR = 5.3, 95% CI: 2.3–12.8, $p < 0.001$), supporting their role as an independent risk factor for recurrence following RT.

Functional analysis revealed a network of interrelated cellular functions involved in actin cytoskeleton regulation, cellular polarity, and membrane dynamics—processes essential for invasive behavior (**Fig. 4d**).

Alterations in *SH2B2* and *SORBS1*, which modulate stress fiber formation and actin remodeling, and in *PTPN13* and *WWC1*, which affect actomyosin tension at the cell leading edge, suggest disruptions in cytoskeletal dynamics. Additional genes associated with prognosis in RT-treated include *KRT1*, *MUCL3*, *TMEM132C*, *TTLL10*, *EML5*, *CEP128*, *CFAP47*, and *KIF24*, all of which contribute to maintaining cellular architecture, polarity, and adhesion (**Fig. 4a**). In addition, alterations in metabolic regulators—such as *MGAM2* and *AADACL3*—could influence the energetic and biosynthetic demands required in response to RT-induced stress, while *REV1* is implicated in DNA damage tolerance, a critical safeguard following genotoxic stress.

Collectively, these findings highlight the pivotal roles of cytoskeletal reorganization, membrane remodeling, metabolic regulation, and DNA repair pathways in the response to RT. The prognostic significance of these mutations in RT-treated patients suggests that these gene alterations may mediate resistance mechanisms triggered by RT-induced stress.

Copy number alterations display molecular subtype-specific patterns with select genomic regions linked to local recurrence

Analysis of copy number alterations (CNAs) revealed recurrent chromosomal changes similar to those reported in invasive breast cancer (**Fig. 5a**). Significant gains were identified on chromosomal arms 1q, 8q, 16p, 17q, 20p, and 20q, while losses predominantly occurred on 8p, 9p, 11q, 13q, 14q, 16q, and 17p (binomial test, FDR < 0.05).

Global CNA burden varied considerably across samples, with approximately half showing minimal alterations (global CNA score < -0.58; **Fig. 5b**). Basal-like tumors exhibited greater

CNA burden, while normal-like tumors showed fewer alterations (**Fig. 5b**, **Supplementary Fig. 2a**). Similarly, high-grade lesions showed a greater CNA burden compared to low-grade lesions (**Supplementary Fig. 2b)**.

We identified distinct CNA patterns across molecular subtypes: basal-like tumors showed enrichment for gains on 8q, 13q, and 19q; LumA tumors frequently exhibited 16q loss; and Her2-enriched tumors showed characteristic 17q12 gains corresponding to the *ERBB2* locus (Fisher's exact test, $p < 0.005$; **Supplementary Fig. 2c**).

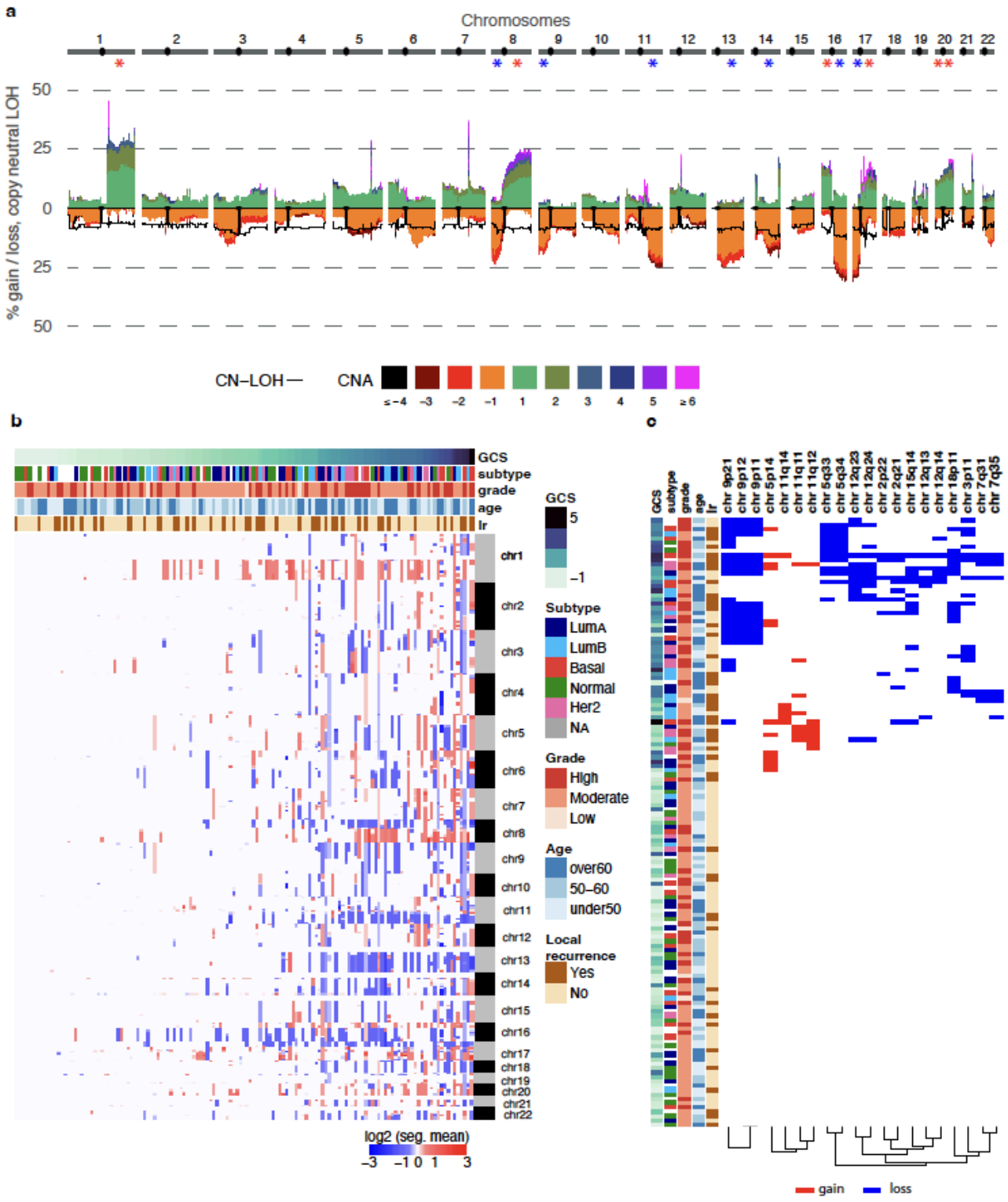

Fig. 5: Frequent copy number alterations in pure DCIS and associations with 10-year LR risk

a Genome-wide frequency of absolute copy number gains and losses across chromosomes 1-22 (1Mb window). The estimated ploidy for each sample is subtracted from the copy number values of each segment which means that a copy number of 0 is no copy number change. The y-axis shows the percentage of samples with each alteration type. Asterisks indicate statistical significance for chromosomal arm alterations **b** Heatmap showing copy number profiles per cytoband across samples (rows), ordered by global CNA score (GCS). Sample annotations include age, grade, molecular subtype and local recurrence status at 10 years. **c** Heatmap of twenty genomic regions significantly associated with 10-year risk of local recurrence (Firth-corrected cox model p-value < 0.05 adjusted for grade), showing copy number status. CNAs present in at least 5 patients are shown, with copy number gains in red and losses in blue. Samples and regions were clustered using Ward's hierarchical clustering with Minkowski distance metric.

Twenty cytoband regions were significantly associated with increased 10-year local recurrence risk after adjustment for grade (Firth's penalized likelihood Cox regression p-value < 0.05; **Fig. 5c**). The alterations predominantly involve copy number losses (12 non-adjacent cytobands) with a smaller number of gains (3 non-adjacent cytobands).

Copy number losses could alter critical tumor suppressor functions, as evidenced by the presence of several well-known tumor suppressor genes within these regions. Notable examples include *CDKN2A* at 9p21 (a key regulator of cell cycle progression), *PRDM4* at 12q23-24 (involved in cell differentiation and tumorigenesis) [28], *MSH2* at 2p21 (a critical component of DNA mismatch repair), *SPARC* at 5q33 (important for extracellular matrix

synthesis and modulation of cell shape) [29], *SPRED1* at 15q14 (which regulates MAPK signaling and epithelial-to-mesenchymal transition) [30], and *ST7* at 7q31 (involved in cell–environment and cell–cell interactions [31]).

Additional losses highlight disruptions in cell adhesion and intercellular contacts. For instance, 12q13-14 harbors a cluster of 28 keratin genes critical for structural integrity and epithelial stability, loss of 18p11 may affect DAL-1, a known regulator of cell adhesion and link between cell membrane and cytoskeleton [32,33], and deletion of 3p11 could impair *EPHA3*, an important regulator of intercellular contacts and cell-cell junction [34].

Gains were identified in regions harboring oncogenes or potential cancer-related genes. For example, amplification in the 11q region—including 11q13, which contains *CCND1* and *CTTN*—previously associated with breast cancer aggressiveness [35,36]. Additionally, a gain in 5p14 includes four cadherin genes (*CDH10*, *CDH9*, *CDH12*, *CDH18*) and *PRDM9*, whose aberrant expression has been associated with genomic instability [37].

Collectively, these findings demonstrate that the CNAs associated with 10-year local recurrence not only affect regions harboring established tumor suppressors and oncogenes but also target regions involved in cell structural integrity and cell-cell interactions, potentially contributing to an aggressive and treatment-resistant phenotype.

Contextualizing our genomic findings with prior profiling studies: consistent early mutational drivers, novel subtype-specific CNA patterns, and predictors of RT response

Most prior DCIS genomic studies analyzed small cohorts (<100 cases) and primarily focused on DCIS cases with concurrent invasive disease (synchronous DCIS) rather than pure DCIS (**Supplementary Table 1**). While recent larger studies, such as those by Strand *et al.* (2022) [20] and Kader et al. (2024) [38], examined hundreds of pure DCIS cases, their reliance on low-pass sequencing or lack of matched normal tissues limits the sensitivity for detecting genomic changes, particularly when working with FFPE samples, which are often the only available option for DCIS studies.

Across studies, *PIK3CA* and *TP53* consistently emerged as the most frequently mutated genes, with mutation rates ranging from 21–55% and 17–52%, respectively (**Supplementary Table 1**). In our cohort, *PIK3CA* and *TP53* were also among the most frequently mutated genes and identified as tumorigenesis drivers in pure DCIS. Mutations in *GATA3* and *PTEN,* which were frequently reported in several studies [13,14,39] , were only detected in a small number of lesions in our cohort (n = 3 for each). Differences across studies likely reflect variations in methodology, including variant-calling pipelines, sample types (e.g., synchronous vs. pure DCIS), cohort composition (e.g., histological grade and ER/HER2 status), and our study's specific focus on LR and RT response, complicating direct comparisons. Nevertheless, the consistent identification of *PIK3CA* and *TP53* as the most frequently mutated genes across studies and in our cohort reinforces their pivotal roles in the early stages of breast tumorigenesis.

Recurrent CNAs in regions such as 1q, 8q, and 17q gains, as well as 8p, 11q, and 16q losses, were observed in our study, consistent with prior DCIS findings, including those from Strand *et al.* (2022) [20] and Abba *et al.* (2015) [15]. For example, Strand *et al.* (2022) identified 29 recurrent CNAs in DCIS but found no single CNA predictive of recurrence [20]. In contrast, we identified twenty genomic regions significantly associated with 10-year local recurrence risk, including regions containing known tumor suppressors and oncogenes. These differences may reflect the higher resolution of our sequencing approach (100x whole-exome sequencing) compared to the low-pass sequencing used in prior studies. Our study also provides valuable new insights into molecular subtype-specific CNA patterns in DCIS. In our cohort, basal-like tumors exhibited gains on 8q, 13q, and 19q; Luminal A tumors displayed 16q loss; and Her2-enriched tumors showed 17q12 gains encompassing *ERBB2*. Aside from amplifications in HER2-positive subtypes, subtype-specific CNA analyses remain poorly represented in the DCIS literature but are well established in invasive breast cancer (**see Supplementary Text**). These findings suggest that genetic and molecular aberrations defining subtypes likely arise early and are at least partially established at the DCIS stage.

Finally, while several studies have explored prognostic markers of recurrence, few account for treatment variation, and none specifically examined markers of RT response (**Supplementary Table 1**). Our study uniquely identified genetic alterations within a gene network that integrates cytoskeletal integrity, cell-cell interactions, cell adhesion, and metabolism which are associated with an increased risk of local recurrence within five years following RT. These findings suggest that these alterations may play a critical role in resistance mechanisms activated by RT-induced stress.

Discussion

Our study provides a comprehensive analysis of pure DCIS, focusing on DNA profiles associated with disease prognosis and response to RT. While previous research has primarily examined genomic changes in synchronous DCIS compared to IDC or pure DCIS, our work offers a detailed view of the mutational landscape specific to pure DCIS and its association with LR. Importantly, we identified genomic alterations associated with molecular subtypes and, most critically, LR—both independent of treatment and specifically in cases following RT—highlighting potential molecular drivers of treatment outcomes. We also observed distinct mutational processes in early-onset DCIS with high tumor mutational burden potentially driven by impaired mismatch repair, but these were not associated with prognosis.

While driver genes like *TP53* and *PIK3CA* are critical in the early stages of tumorigenesis, our study did not find any significant associations between their mutations and disease progression or response to treatment. These findings suggest that TP53 and PIK3CA mutations are essential for overcoming initial biological constraints during early tumor development but may become less relevant to progression once these barriers are surpassed. This aligns with prior research showing that these mutations are frequently retained in invasive carcinoma but occur at higher prevalence in hyperplastic and DCIS lesions [40].

Furthermore, their prognostic relevance appears to depend on tumor clinicopathological characteristics. For example, Lin et al. [11] observed an inverse association between *PIK3CA* kinase domain mutations in high-grade DCIS tumors and progression, while Silwal-Pandit et al. [41] reported that the prognostic impact of *TP53* mutations varied across molecular subtypes of invasive breast cancer. These findings underscore the complexity of breast cancer

progression, where early driver mutations interact with additional genetic and microenvironmental factors to shape disease outcomes.

We identified recurrent CNAs frequently reported in the literature, including gains on 8q [10–12,14,15,38,39], 17q [10,12,15,18–20,38,39], 20q [38,39], and losses on 11q [10–12,18,19,38] and 16q [10,13,14,20,39]. These CNAs reflect patterns similar to those observed in IDC, suggesting that pure DCIS already harbors genomic features characteristic of invasive cancer including CNAs defining molecular subtypes such as basal, luminal-A or Her2-enriched tumors (see **Supplementary Text**).

Importantly, our study also revealed novel associations between mutations in genes regulating cytoskeletal integrity, cell membrane organization and function, cell-cell interactions, and cell adhesion, and the risk of invasive or in situ LR in pure DCIS. Across the cohort, five genes were significantly associated with LR, underscoring their potential roles in tumor progression. These mutations were mostly mutually exclusive, not frequently co-occurring in the same patient, and involved genes maintaining cell adhesion (*MYO7A*, *MUCL3*), organizing and supporting cellular membrane function (*SH2B2*, *PDZD8*), and facilitating DNA degradation (*DNASE2B*). When the cohort was stratified by RT administration, 27 unique genes - excluding *SH2B2* and *MUCL3* - were significantly associated with increased LR risk in RT-treated patients. Many of these mutations were identified in genes critical for maintaining cellular architecture, polarity, and adhesion, such as *KRT1*, *MUCL3*, *TMEM132C*, *TTLL10*, *EML5*, *CEP128*, *CFAP47*, and *KIF24*, highlighting how disruptions in structural integrity and epithelial cell organization may undermine the effectiveness of RT and contribute to LR.

In addition to structural changes, metabolic regulators like *MGAM2* and *AADACL3* may influence the energetic and biosynthetic demands required for cell survival and adaptation under genotoxic conditions. Alterations in *REV1*, a key player in DNA damage tolerance, underscore the importance of effective DNA repair mechanisms following RT. These findings expand upon prior studies linking disruptions in tissue structure and cell adhesion to DCIS progression [42–45], while providing novel insights into their specific roles in recurrence following RT in pure DCIS.

Finally, we uniquely identified losses in cytobands harboring established tumor suppressor genes—such as *CDKN2A* in 9p21, *PRDM4* in 12q23–24, and *MSH2* in 2p21—as well as regions containing genes crucial for cell adhesion and migration, including a cluster of keratin genes in 12q13–14, *DAL-1* in 18p11, and *EPHA3* in 3p11. In parallel, gains in regions like 11q, which includes oncogenic drivers such as *CCND1* and *CTTN*, suggest that oncogene amplification may further contribute to tumor aggressiveness and ultimately patient prognosis. Gains at 5p14, containing cadherin genes (*CDH10*, *CDH9*, *CDH12*, *CDH18*), again emphasize the importance of maintaining epithelial integrity to prevent LR in pure DCIS patients.

Future studies are warranted to investigate how these mutations interact with other molecular pathways and microenvironmental factors to elucidate their contribution to DCIS prognosis and the adverse effects of RT. Such research could help identify strategies to mitigate recurrence risk following RT and improve treatment outcomes [46–48].

Conclusions

Our findings uncover the genomic landscape of pure DCIS and highlight key factors that contribute to LR and mediate the adverse effects of RT. While *TP53* and *PIK3CA* mutations play important roles in early tumorigenesis, they do not predict recurrence, emphasizing the need for alternative biomarkers. We identify distinct mutational processes and genetic alterations that disrupt cytoskeletal integrity, cell-cell interactions and cell adhesion, potentially destabilizing the epithelial tissue environment and contributing to recurrence, particularly following RT-induced stress. These findings provide a foundation for understanding the genetic basis of DCIS progression and identifying potential molecular drivers of treatment resistance. Future research will be essential to translate these insights into clinical practice, guiding the development of more targeted therapeutic approaches to improve outcomes for patients with DCIS.

Methods

The Ontario DCIS cohort

The Ontario DCIS Cohort was established at the Sunnybrook Health Sciences Center (Toronto, Canada) as a population-based sample of women diagnosed with pure DCIS defined as in situ cancer without any invasive component between 1994 and 2003 [49–51]. All patients underwent BCS, with a subset receiving subsequent RT. Adjuvant endocrine therapy was administered to less than 15% of individuals, while none received systemic chemotherapy or neoadjuvant endocrine therapy. The cohort features comprehensive annotation of clinical annotation and expert pathology review. Previous studies of this cohort have characterized

outcomes based on clinical factors including age at diagnosis, pathological features (tumor size and nuclear grade), and treatment modalities [52–55].

Sample selection was prioritized to achieve balanced representation between RT-treated and untreated patients, as well as between patients who experienced invasive or in-situ ipsilateral LR within ten years post-treatment and those who did not. Tissue cores were obtained from FFPE blocks, sampling DCIS tumors without microinvasion alongside adjacent normal and stromal tissues. DNA and RNA were extracted using the Qiagen AllPrep FFPE DNA/RNA kit (Qiagen). Samples yielding sufficient DNA quantities underwent library construction using the Nextera Flex RNA Exome kit (Illumina) and were sequenced on the NovaSeq6000 platform (100bp paired-end, 100M reads/sample) at the Genome Quebec Innovation Centre (Montreal, Canada). While high-quality sequencing data was obtained for 300 tumor tissues, downstream analyses focused on 147 samples with matched normal profiles (144 normal tissue and 3 stroma non-epithelial samples).

Molecular subtypes were determined using RNA profiles available for a subset of patients (n = 122). PAM50 subtype classification was performed using the `genefu` R package with established centroids. Specifically, normalized expression data of the 50 PAM50 genes were obtained using variance stabilizing transformation implemented in the `DESeq2` R package [56] and compared to subtype-specific centroids using Pearson correlations. Each sample was assigned to the molecular subtype with which it showed the highest correlation coefficient.

Whole-exome sequencing data preprocessing

Raw reads were processed using Trimmomatic [57] (version 0.39) to remove adaptor sequences and low quality bases. Reads were trimmed to retain high-quality sequences, applying quality thresholds of 10 at read ends and 20 within a 4-base sliding window. The remaining paired reads were processed following the GATK4 best practices. Briefly, reads were aligned to the human reference genome (GRCh38, GATK resource bundle) using the Burrows-Wheeler Aligner (BWA) [58]. Post-alignment procedures included sorting, annotating reads with read groups, and marking duplicate reads with Picard. Base quality score recalibration was conducted using GATK4 tools. A recalibration table was generated with the `BaseRecalibrator` function using known variant sites (dbSNP138 for SNPs and Mills and 1000 Genomes for indels), and recalibration was applied with `ApplyBQSR` to adjust base quality scores and correct for systematic technical errors. The process focused on SureSelect Human Exons v7 regions with a 100 bp padding. Finally, properly paired reads were extracted, excluding secondary alignments and low-quality reads, with the resulting files indexed using `Samtools` [59].

Single nucleotide variant & indel calling

For variant calling, we used NeuSomatic [60], a deep learning approach that leverages both tumor and matched normal sequence alignment information, alongside somatic mutation calls from six different approaches: MuTect2, MuSE, VarDict, VarScan2, Strelka2, and SomaticSniper [61–66]. This method was selected because of the low level of agreement between callers in our data - the majority of mutations (88.7%) were identified by only one caller (**Supplementary Fig. 3A**) - an observation consistent with previous studies [67–69]. The number of mutations

detected varied significantly across samples, with the minimum identified in any single sample being 1,563 mutations, and the maximum reaching 283,247 mutations (**Supplementary Fig. 3A**).

We used the ensemble extension of NeuSomatic which includes 93 channels to capture features extracted from the six individual methods and 26 additional channels to capture the alignment information in a window of seven bases around the candidate mutation. This results in 119 input channels for each candidate matrix. We used the recommended pre-trained model SEQC-II (SEQC-WGS-Spike model) (trained on 20 whole-genome sequencing replicate pairs with *in silico* somatic mutations of 1%-100% AF, matched with both 95%N and 100%N, ~40x-220x, 5 callers used: MuTect2, Strelka2, MuSE, SomaticSniper, VarDict). After mutation calling, the recommended post-processing was applied to resolve long INDEL sequences. The final NeuSomatic predictions were used for downstream analyses.

We obtained comprehensive genomic information for each variant using the Ensembl Variant Effect Predictor (VEP) [70]. This included the effects on gene and protein function, such as consequence types and amino acid changes, variant frequencies in different populations, impact on regulatory regions, and potential associations with diseases and phenotypes. Following annotation with the Ensembl VEP, we used the `vcf2maf` tool to transform VEP-annotated VCF files into the Mutation Annotation Format (MAF). This conversion ensures each variant is uniquely associated with a single gene transcript or isoform, despite the potential for a variant to impact multiple isoforms. Particularly in cases where variants could be classified under different effects — such as a `Missense_Mutation` near a `Splice_Site`

— the MAF format forces a singular designation for each variant by leveraging VEP's determinations for canonical isoforms.

We excluded 100 genes commonly mutated in public exome datasets (FLAGS) due to their lower likelihood of disease association [71]. This decision stems from their longer coding regions, which inherently increase mutation probability, and the presence of paralogs that might offset functional loss these mutations could cause [71,72].

High-confidence variants (identified with a probability score of 0.7 or higher) consistently showed higher allele frequencies compared to those categorized as low-quality (with scores between 0.4 and 0.7) and rejected variants (with scores below 0.4) (**Supplementary Fig. 3B**). To reduce potential false positives, we selected high-confidence variants with allele frequency above 0.1.

Mutation patterns and frequencies were visualized using the `oncoplot` function from the `maftools` R package [73], which displays mutation types and frequencies across samples.

Mutational signatures

We performed mutational signature analysis using the COSMIC database of single-base substitution (SBS) signatures [24]. First, the trinucleotide context of each single nucleotide variant was characterized. We then used the `fit_to_signature` function in the `MutationalPatterns` R package (version 3.19) to find the linear combination of COSMIC mutation signatures that most closely reconstructs the mutation spectra for each sample by solving the nonnegative least-squares constraints problem. We used strict refit where the

signature with the lowest contribution is removed; refitting is repeated until the cosine similarity between the original and reconstructed profile becomes more than max_delta= 0.004). We selected signatures that contributed to the mutation spectrum of at least 10 samples and plotted their relative contributions using the `pheatmap` R package.

Driver genes

To identify driver genes (i.e., genes under positive selection in cancer), we used the `dNdScv` analysis method [74]. This approach is based on the evaluation of the ratio between synonymous (silent) mutations and non-synonymous (missense) mutations in genes. A higher ratio of non-synonymous to synonymous mutations in a gene indicates positive selection for mutations that may confer a growth advantage to cancer cells, suggesting the potential role of the gene as a driver in tumorigenesis.

The `dNdScv` method estimates the background mutation rate of each gene by combining local information (synonymous mutations within the gene) with global information (variation of mutation rates across genes). This approach controls for the sequence composition of genes and accounts for mutational signatures, providing a more accurate estimation of the expected neutral mutation rate. In particular, the `dNdScv` R package implementation uses trinucleotide context-dependent substitution matrices to mitigate common mutation biases that can affect dN/dS calculations [74].

To visualize and analyze the distribution and nature of mutations in driver genes, we used lollipop plots generated from the maftools R package and annotation tracks from cBioPortal. These plots provide a representation of mutation types and their locations along the protein sequence as well as annotations ncluding likely mutation hotspots as identified by Memorial

Sloan Kettering Cancer Hotspots and 3D Hotspots databases [77], and annotation records of therapeutic indication from OncoKB [78].

Copy Number Alterations (CNAs)

To investigate copy number alterations, we applied the Allele-Specific Copy Number Analysis of Tumours (ASCAT v3) on our tumor normal pairs estimating tumor purity, ploidy, and allele-specific copy number [80]. The `runAscat` function was executed with default settings optimized for high-throughput exome sequencing data, with the gamma parameter set to 1. After examination of ASCAT sunrise plots, we identified a subset of samples (n=35 samples), for which the initial estimates of tumor purity and ploidy did not align with the regions of highest confidence on the sunrise plots. The `runAscat` was re-run for these samples by manually assigning the aberrant cell fraction (tumor purity) and tumor ploidy parameters corresponding to the regions of highest probability as depicted on the sunrise plots. Overall, eight samples were excluded from further analysis due to poor goodness of fit leaving 139 samples with CNA profiles for downstream analyses. Absolute number gains and losses shared across samples were visualized across whole chromosomal regions using `aCNViewer` [81] (window size of 1 Mbp). The estimated ploidy for each sample is subtracted from the copy number values of each segment which means that a copy number of 0 is copy number change. These adjusted windows at base resolution are then plotted into a stacked histogram representing genome-wide absolute copy number and copy neutral variations over all samples in a group.

We applied a re-segmentation approach to adjust for amplitude divergence due to technical variability implemented in `CNApp` [82] using the default settings (minimum segment length = 100 Kbp, minimum amplitude deviation from segment to zero = 0.16, maximum distance between segments=1 Mb, maximum amplitude deviation between segments = 0.16, and maximum BAF deviation between segments = 0.1). Re-segmented data were then used to calculate the broad, focal and global CNA scores. We then transformed re-segmented data into genomic regions profiles (chromosome arms, cytobands and sub-cytobands) using both focal and broad segments. Length-relative means are computed for each window by considering amplitude values from those segments included in each specific window. Default cutoffs for low-level copy number gains and losses (i.e., |0.2|) were used to infer CNA frequencies.

Survival analyses

We evaluated the association between gene mutations or copy number aberrations in cytobands and 10-year local recurrence-free survival using Firth's penalized likelihood Cox regression which accounts for small sample sizes and rare events. This analysis was conducted using the `coxphf` R package. Aberrations were included in the analysis only if detected in at least five lesions. This analysis was performed across the entire patient cohort. To further investigate genetic alterations associated with response to RT, a stratified analysis was conducted based on treatment groups.

Kaplan-Meier survival curves were used to visualize the results, illustrating event-free survival probabilities over time for patients stratified by mutational status in the specified genes or gene sets.

GO enrichment analyses were performed using the clusterProfiler R package. Entrez gene identifiers were mapped to GO terms using the `org.Hs.eg.db` annotation database. All GO terms with FDR < 0.2 were considered. Semantic similarity between GO terms was calculated using the Wang method implemented in the `pairwise_termsim` function. The enrichment map was visualized using the `emapplot` function which displays the significantly enriched terms with at least 2 genes.

Third-party studies

A systematic literature search was performed using PubMed to identify previous studies that conducted DNA profiling on pure DCIS or DCIS mixed with invasive lesions. The search strategy included terms related to “ductal carcinoma in situ”, “genetic markers”, “DCIS prognosis”, “DCIS progression”, “DCIS to IDC”, "dcis dna", "dcis prognosis dna markers", "copy number alterations", and “somatic mutations”. Studies were included if they reported genomic analyses of DCIS samples using sequencing or copy number profiling techniques and were published within the last 10 years. Twelve studies met the inclusion criteria, and their key findings were summarized in **Supplementary Table 1**. The review emphasized genetic alterations and pathway dysregulation that may drive DCIS initiation and progression to invasive disease.

Declarations

Ethics approval and consent to participate

Research ethics board approval was obtained from Sunnybrook Health Sciences Centre, Toronto, Ontario. This study was facilitated through ICES which is named as a prescribed entity in section 45 of PHIPA (Regulation 329/04, section 18) which allows access and utilization of administrative data for research purposes with a waived requirement for consent.

Consent for publication

Not Applicable

Availability of data and materials

The dataset analysed during the current study is not publicly available due the personal, sensitive, and inherently identifying nature of raw genomic data. Access to raw data and patient metadata may be provided but is controlled and requires institutional material data transfer agreements (contact person: eileen.rakovitch@sunnybrook.ca).

The scripts used to reproduce the analyses performed in this study are available in the Dumeaux Lab GitHub Repository https://github.com/dumeaux-lab/dcis-dna_paper.

Competing Interests

The authors declare no competing interests.

Funding

This work was supported by Canadian Institute for Health Research (CIHR) project grant #391682. We thank the Canadian Foundation for Innovation J. Evans Leaders Fund grant #43481 for the support of the computing infrastructure used throughout our analyses.

ER holds LC Campbell Chair for breast cancer research. NR received Breast Cancer Canada and Ontario Graduate Scholarships.

We thank Genome Quebec Innovation Center for RNA processing and sequencing.

Authors' Contributions

Conceptualization, S.N.M, L.P., E.R., M.T.H and V.D.; Methodology, N.R, E.S, V.D.; Investigation, N.R., E.S., M.Z., V.D.; Software N.R., E.J.M., V.D.; Writing – Original Draft, N.R. and V.D.; Writing – Review & Editing, E.S., M.Z., E.J.M., E.H., S.N.M., S.T., L.P., M.T.H. and E.R. ; Funding Acquisition, E.R., M.T.H, E.H., S.N.M, S.T., L.P. and V.D.; Resources, E.R., S.N.M, S.T., L.P. ; Supervision, V.D.

References

1. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nat Rev Dis Primer. 2019;5:1–31.

2. Sørum R, Hofvind S, Skaane P, Haldorsen T. Trends in incidence of ductal carcinoma in situ: The effect of a population-based screening programme. The Breast. 2010;19:499–505.

3. Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjær H, Holford TR, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2015;24:905–12.

4. Force UPST, Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, et al. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2024;331:1918–30.

5. Tomlinson-Hansen S, Khan M, Cassaro S. Breast Ductal Carcinoma in Situ. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 [cited 2023 Nov 16]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK567766/

6. Zhang Y, Weinberg RA. Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. Front Med. 2018;12:361–73.

7. Maxwell AJ, Hilton B, Clements K, Dodwell D, Dulson-Cox J, Kearins O, et al. Unresected screen-detected ductal carcinoma in situ: Outcomes of 311 women in the Forget-Me-Not 2 study. Breast Edinb Scotl. 2022;61:145–55.

8. Rakovitch E, Gray R, Baehner FL, Sutradhar R, Crager M, Gu S, et al. Refined estimates of local recurrence risks by DCIS score adjusting for clinicopathological features: a combined analysis of ECOG-ACRIN E5194 and Ontario DCIS cohort studies. Breast Cancer Res Treat. 2018;169:359–69.

9. Andor N, Maley CC, Ji HP. Genomic instability in cancer: Teetering on the limit of tolerance. Cancer Res. 2017;77:2179–85.

10. Trinh A, Gil Del Alcazar CR, Shukla SA, Chin K, Chang YH, Thibault G, et al. Genomic Alterations during the In Situ to Invasive Ductal Breast Carcinoma Transition Shaped by the Immune System. Mol Cancer Res. 2021;19:623–35.

11. Lin C-Y, Vennam S, Purington N, Lin E, Varma S, Han S, et al. Genomic landscape of ductal carcinoma in situ and association with progression. Breast Cancer Res Treat. 2019;178:307–16.

12. Hernandez L, Wilkerson PM, Lambros MB, Campion-Flora A, Rodrigues DN, Gauthier A, et al. Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. J Pathol. 2012;227:42–52.

13. Nagasawa S, Kuze Y, Maeda I, Kojima Y, Motoyoshi A, Onishi T, et al. Genomic profiling reveals heterogeneous populations of ductal carcinoma in situ of the breast. Commun Biol. 2021;4:1–13.

14. Nachmanson D, Officer A, Mori H, Gordon J, Evans MF, Steward J, et al. The breast pre-cancer atlas illustrates the molecular and micro-environmental diversity of ductal carcinoma in situ. Npj Breast Cancer. 2022;8:1–13.

15. Abba MC, Gong T, Lu Y, Lee J, Zhong Y, Lacunza E, et al. A Molecular Portrait of High-Grade Ductal Carcinoma *In Situ*. Cancer Res. 2015;75:3980–90.

16. Kaplan HG, Dowdell AK, Berry AB, Shimol RB, Robinson FL, Carney CA, et al. Multi-omic profiling of simultaneous ductal carcinoma in situ and invasive breast cancer. Breast Cancer Res Treat. 2024;205:451–64.

17. Kader T, Hill P, Zethoven M, Goode DL, Elder K, Thio N, et al. Atypical ductal hyperplasia is a multipotent precursor of breast carcinoma. J Pathol. 2019;path.5262.

18. Kim SY, Jung S-H, Kim MS, Baek I-P, Lee SH, Kim T-M, et al. Genomic differences between pure ductal carcinoma in situ and synchronous ductal carcinoma in situ with invasive breast cancer. Oncotarget. 2015;6:7597–607.

19. Pareja F, Brown DN, Lee JY, Da Cruz Paula A, Selenica P, Bi R, et al. Whole-Exome Sequencing Analysis of the Progression from Non–Low-Grade Ductal Carcinoma *In Situ* to Invasive Ductal Carcinoma. Clin Cancer Res. 2020;1078-0432.CCR-19–2563.

20. Strand SH, Rivero-Gutiérrez B, Houlahan KE, Seoane JA, King LM, Risom T, et al. Molecular classification and biomarkers of clinical outcome in breast ductal carcinoma in situ: Analysis of TBCRC 038 and RAHBT cohorts. Cancer Cell [Internet]. 2022 [cited 2022 Dec 7]; Available from: https://www.sciencedirect.com/science/article/pii/S1535610822005128

21. Cooper DN, Mort M, Stenson PD, Ball EV, Chuzhanova NA. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. Hum Genomics. 2010;4:406–10.

22. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47:1402–7.

23. Steiert TA, Parra G, Gut M, Arnold N, Trotta J-R, Tonda R, et al. A critical spotlight on the paradigms of FFPE-DNA sequencing. Nucleic Acids Res. 2023;51:7143–62.

24. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.

25. Guo Q, Lakatos E, Bakir IA, Curtius K, Graham TA, Mustonen V. The mutational signatures of formalin fixation on the human genome [Internet]. bioRxiv; 2021 [cited 2024 Dec 6]. p. 2021.03.11.434918. Available from: https://www.biorxiv.org/content/10.1101/2021.03.11.434918v1

26. Yuan T, Cantley L. PI3K pathway alterations in cancer: variations on a theme. Oncogene. 2008;27:5497–510.

27. Bader AG, Kang S, Zhao L, Vogt PK. Oncogenic PI3K deregulates transcription and translation. Nat Rev Cancer. 2005;5:921–9.

28. Yang XH, Huang S. PFM1 (PRDM4), a new member of the PR-domain family, maps to a tumor suppressor locus on human chromosome 12q23-q24.1. Genomics. 1999;61:319–25.

29. Arnold SA, Brekken RA. SPARC: a matricellular regulator of tumorigenesis. J Cell Commun Signal. 2009;3:255–73.

30. Lorenzo C, McCormick F. SPRED proteins and their roles in signal transduction, development, and malignancy. Genes Dev. 2020;34:1410–21.

31. Zenklusen JC, Conti CJ, Green ED. Mutational and functional analyses reveal that ST7 is a highly conserved tumor-suppressor gene on human chromosome 7q31. Nat Genet. 2001;27:392–8.

32. Heller G, Geradts J, Ziegler B, Newsham I, Filipits M, Markis-Ritzinger E-M, et al. Downregulation of TSLC1 and DAL-1 expression occurs frequently in breast cancer. Breast Cancer Res Treat. 2007;103:283–91.

33. Kittiniyom K, Gorse KM, Dalbegue F, Lichy JH, Taubenberger JK, Newsham IF. Allelic loss on chromosome band 18p11.3 occurs early and reveals heterogeneity in breast cancer progression. Breast Cancer Res. 2001;3:192–8.

34. Pasquale EB. Eph receptors and ephrins in cancer: bidirectional signaling and beyond. Nat Rev Cancer. 2010;10:165–80.

35. Schuuring E, Verhoeven E, Mooi WJ, Michalides RJ. Identification and cloning of two overexpressed genes, U21B31/PRAD1 and EMS1, within the amplified chromosome 11q13 region in human carcinomas. Oncogene. 1992;7:355–61.

36. Ormandy CJ, Musgrove EA, Hui R, Daly RJ, Sutherland RL. Cyclin D1, EMS1 and 11q13 Amplification in Breast Cancer. Breast Cancer Res Treat. 2003;78:323–35.

37. Houle AA, Gibling H, Lamaze FC, Edgington HA, Soave D, Fave M-J, et al. Aberrant PRDM9 expression impacts the pan-cancer genomic landscape. Genome Res. 2018;28:1611–20.

38. Kader T, Zethoven M, Mahale S, Saunders H, Tjoeka L, Lehmann R, et al. Predictive biomarkers of breast ductal carcinoma in situ may underestimate the risk of recurrence due to de novo ipsilateral breast carcinoma development [Internet]. bioRxiv; 2024 [cited 2025 Jan 3]. p. 2024.05.19.594731. Available from: https://www.biorxiv.org/content/10.1101/2024.05.19.594731v1

39. Pang J-MB, Savas P, Fellowes AP, Mir Arnau G, Kader T, Vedururu R, et al. Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. Mod Pathol. 2017;30:952–63.

40. Ang DC, Warrick AL, Shilling A, Beadling C, Corless CL, Troxell ML. Frequent phosphatidylinositol-3-kinase mutations in proliferative breast lesions. Mod Pathol. 2014;27:740–50.

41. Silwal-Pandit L, Langerød A, Børresen-Dale A-L. TP53 Mutations in Breast and Ovarian Cancer. Cold Spring Harb Perspect Med. 2017;7:a026252.

42. Emery LA, Tripathi A, King C, Kavanah M, Mendez J, Stone MD, et al. Early Dysregulation of Cell Adhesion and Extracellular Matrix Pathways in Breast Cancer Progression. Am J Pathol. 2009;175:1292–302.

43. Glukhova MA, Streuli CH. How integrins control breast biology. Curr Opin Cell Biol. 2013;25:633–41.

44. Maxwell CA, Benítez J, Gómez-Baldó L, Osorio A, Bonifaci N, Fernández-Ramires R, et al. Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer. PLoS Biol. 2011;9:e1001199.

45. Qin X, Strand SH, Lee MR, Saraswathibhatla A, van IJzendoorn DGP, Zhu C, et al. Single Cell Expression Analysis of Ductal Carcinoma in Situ Identifies Complex Genotypic-Phenotypic

Relationships Altering Epithelial Composition. Cancer Res [Internet]. 2025 [cited 2025 Apr 21]; Available from: https://doi.org/10.1158/0008-5472.CAN-24-3023

46. Deng Y, Chen G, Xiao J, Deng H. Role and potential therapeutic strategies of matrix mechanics for optimizing tumor radiotherapy. Mechanobiol Med. 2024;2:100037.

47. Prakash J, Shaked Y. The Interplay between Extracellular Matrix Remodeling and Cancer Therapeutics. Cancer Discov. 2024;14:1375–88.

48. Frascogna C, Mottareale R, La Verde G, Arrichiello C, Muto P, Netti PA, et al. Role of the mechanical microenvironment on CD-44 expression of breast adenocarcinoma in response to radiotherapy. Sci Rep. 2024;14:391.

49. Rakovitch E, Nofech-Mozes S, Narod SA, Hanna W, Thiruchelvam D, Saskin R, et al. Can we select individuals with low risk ductal carcinoma in situ (DCIS)? A population-based outcomes analysis. Breast Cancer Res Treat. 2013;138:581–90.

50. Rakovitch E, Nofech-Mozes S, Hanna W, Baehner FL, Saskin R, Butler SM, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. Breast Cancer Res Treat. 2015;152:389–98.

51. Rakovitch E, Nofech-Mozes S, Hanna W, Sutradhar R, Baehner FL, Miller DP, et al. Multigene Expression Assay and Benefit of Radiotherapy After Breast Conservation in Ductal Carcinoma in Situ. J Natl Cancer Inst. 2017;109:djw256.

52. Paszat L, Sutradhar R, Zhou L, Nofech-Mozes S, Rakovitch E. Including the Ductal Carcinoma-In-Situ (DCIS) Score in the Development of a Multivariable Prediction Model for Recurrence After Excision of DCIS. Clin Breast Cancer. 2019;19:35–46.

53. Klein J, Kong I, Paszat L, Nofech-Mozes S, Hanna W, Thiruchelvam D, et al. Close or positive resection margins are not associated with an increased risk of chest wall recurrence in women with DCIS treated by mastectomy: a population-based analysis. SpringerPlus. 2015;4:335.

54. Lalani N, Paszat L, Sutradhar R, Thiruchelvam D, Nofech-Mozes S, Hanna W, et al. Long-term outcomes of hypofractionation versus conventional radiation therapy after breast-conserving surgery for ductal carcinoma in situ of the breast. Int J Radiat Oncol Biol Phys. 2014;90:1017–24.

55. Rakovitch E, Parpia S, Koch A, Grimard L, Soliman H, Stevens C, et al. DUCHESS: an evaluation of the ductal carcinoma in situ score for decisions on radiotherapy in patients with low/intermediate-risk DCIS. Breast Cancer Res Treat. 2021;188:133–9.

56. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.

57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014;30:2114–20.

58. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv.org. 2013 [cited 2024 Apr 8]. Available from: https://arxiv.org/abs/1303.3997v2

59. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10:giab008.

60. Sahraeian SME, Liu R, Lau B, Podesta K, Mohiyuddin M, Lam HYK. Deep convolutional neural networks for accurate somatic mutation detection. Nat Commun. 2019;10:1041.

61. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2 [Internet]. 2019 [cited 2024 Apr 8]. Available from: http://biorxiv.org/lookup/doi/10.1101/861054

62. Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal PA, et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol. 2016;17:178.

63. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016;44:e108.

64. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

65. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15:591–4.

66. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics. 2012;28:311.

67. Sahraeian SME, Fang LT, Karagiannis K, Moos M, Smith S, Santana-Quintero L, et al. Achieving robust somatic mutation detection with deep learning models derived from reference data sets of a cancer sample. Genome Biol. 2022;23:12.

68. Barbitoff YA, Abasov R, Tvorogova VE, Glotov AS, Predeus AV. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. BMC Genomics. 2022;23:155.

69. de Schaetzen van Brienen L, Larmuseau M, Van der Eecken K, De Ryck F, Robbe P, Schuh A, et al. Comparative analysis of somatic variant calling on matched FF and FFPE WGS samples. BMC Med Genomics. 2020;13:94.

70. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.

71. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJ, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. BMC Med Genomics. 2014;7:64.

72. Alfieri F, Caravagna G, Schaefer MH. Cancer genomes tolerate deleterious coding mutations through somatic copy number amplifications of wild-type regions. Nat Commun. 2023;14:3594.

73. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747–56.

74. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Loo PV, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041.e21.

75. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11:863–74.

76. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.

77. Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P, et al. Accelerating Discovery of Functional Mutant Alleles in Cancer. Cancer Discov. 2018;8:174–83.

78. Suehnholz SP, Nissan MH, Zhang H, Kundra R, Nandakumar S, Lu C, et al. Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. Cancer Discov. 2024;14:49–65.

79. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017;49:170–4.

80. Lesluyes T, Tarabichi M, Haase K, Demeulemeester J, Loo P. Robust and platform-independent CNA calling with ASCAT v3. 2023.

81. Renault V, Tost J, Pichon F, Wang-Renault S-F, Letouzé E, Imbeaud S, et al. aCNViewer: Comprehensive genome-wide visualization of absolute copy number and copy neutral variations. PloS One. 2017;12:e0189334.

82. Franch-Expósito S, Bassaganyas L, Vila-Casadesús M, Hernández-Illán E, Esteban-Fabró R, Díaz-Gay M, et al. CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. Robles-Espinoza CD, Cheah KSE, editors. eLife. 2020;9:e50267.