Modern Pathology (2017) 30, 952–963
Official journal of the United States and Canadian Academy of Pathology

Breast ductal carcinoma *in situ* carry mutational driver events representative of invasive breast cancer

Jia-Min B Pang[1,2], Peter Savas[3], Andrew P Fellowes[1], Gisela Mir Arnau[4], Tanjina Kader[5,6], Ravikiran Vedururu[1], Chelsee Hewitt[1], Elena A Takano[1], David J Byrne[1], David YH Choong[1], Ewan KA Millar[7,8,9,10], C Soon Lee[10,11,12], Sandra A O'Toole[7,11,13], Sunil R Lakhani[14,15], Margaret C Cummings[14,15], G Bruce Mann[16,17], Ian G Campbell[5,6], Alexander Dobrovic[18], Sherene Loi[3,5], Kylie L Gorringe[2,5,6,19] and Stephen B Fox[1,2,5,19]

[1]Department of Pathology, Peter MacCallum Cancer Centre, Grattan Street, Melbourne, VIC, Australia; [2]Department of Pathology, University of Melbourne, Grattan Street, Parkville, Melbourne, VIC, Australia; [3]Division of Research and Cancer Medicine, Peter MacCallum Cancer Centre, Grattan Street, Melbourne, VIC, Australia; [4]Molecular Genomics Core, Peter MacCallum Cancer Centre, Grattan Street, Melbourne, VIC, Australia; [5]Sir Peter MacCallum Department of Oncology, University of Melbourne, Grattan Street, Parkville, Melbourne, VIC, Australia; [6]Cancer Genomics Program, Peter MacCallum Cancer Centre, Grattan Street, Melbourne, VIC, Australia; [7]Translational Breast Cancer Research, The Kinghorn Cancer Centre and Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; [8]Department of Anatomical Pathology, South Eastern Area Pathology Service, St George Hospital, Kogarah, NSW, Australia; [9]School of Medical Sciences, University of New South Wales, Kensington, NSW, Australia; [10]Discipline of Pathology, School of Medicine, University of Western Sydney, Campbelltown, NSW, Australia; [11]Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, NSW, Australia; [12]Cancer Pathology, Bosch Institute, University of Sydney, Camperdown, NSW, Australia; [13]Sydney Medical School, University of Sydney, Camperdown, NSW, Australia; [14]School of Medicine and UQ Centre for Clinical Research, The University of Queensland, Brisbane, Australia; [15]Pathology Queensland, The Royal Brisbane and Women's Hospital, Brisbane, Australia; [16]Breast Service, The Royal Melbourne and Royal Women's Hospitals, Parkville, VIC, Australia; [17]Department of Surgery, University of Melbourne, Grattan Street, Parkville, Melbourne, VIC, Australia and [18]Translational Genomics & Epigenomics Laboratory, Olivia Newton-John Cancer Research Institute, Heidelberg, VIC, Australia

The spectrum of genomic alterations in ductal carcinoma *in situ* (DCIS) is relatively unexplored, but is likely to provide useful insights into its biology, its progression to invasive carcinoma and the risk of recurrence. DCIS ($n = 20$) with a range of phenotypes was assessed by massively parallel sequencing for mutations and copy number alterations and variants validated by Sanger sequencing. *PIK3CA* mutations were identified in 11/20 (55%), *TP53* mutations in 6/20 (30%), and *GATA3* mutations in 9/20 (45%). Screening an additional 91 cases for *GATA3* mutations identified a final frequency of 27% (30/111), with a high proportion of missense variants (8/30). *TP53* mutations were exclusive to high grade DCIS and more frequent in PR-negative tumors compared with PR-positive tumors ($P = 0.037$). *TP53* mutant tumors also had a significantly higher fraction of the genome altered by copy number than wild-type tumors ($P = 0.005$), including a significant positive association with amplification or gain of *ERBB2* ($P < 0.05$). The association between *TP53* mutation and *ERBB2* amplification was confirmed in a wider DCIS cohort using p53 immunohistochemistry as a surrogate marker for *TP53* mutations ($P = 0.03$). *RUNX1* mutations and *MAP2K4* copy number loss were novel findings in DCIS. Frequent copy number alterations included gains on 1q, 8q, 17q, and 20q and losses on 8p, 11q, 16q, and 17p. Patterns of genomic alterations observed in DCIS were similar to those previously reported for invasive breast cancers, with all DCIS having at least one bona fide breast cancer driver event. However, an increase in *GATA3* mutations and fewer copy number changes were noted in DCIS compared with invasive carcinomas. The role of such alterations as prognostic and predictive biomarkers in DCIS is an avenue for further investigation.
Modern Pathology (2017) **30,** 952–963; doi:10.1038/modpathol.2017.21; published online 24 March 2017

Correspondence: Dr KL Gorringe, PhD, Cancer Genomics Program, Peter MacCallum Cancer Centre, Grattan Street, Melbourne, VIC 3000, Australia.
E-mail: kylie.gorringe@petermac.org
[19]These authors contributed equally to this work.
Received 5 January 2017; revised 11 February 2017; accepted 19 February 2017; published online 24 March 2017

The mutational landscape of invasive breast carcinoma has been extensively documented in recent landmark studies.[1–5] However, little is known about the mutational profile of ductal carcinoma *in situ* (DCIS). Studies have addressed mutations of individual genes[6–13] but until very recently, assessment of a broad panel of genes in DCIS has not been performed, largely due the challenges in obtaining DNA from DCIS cases compatible with highly multiplexed methodologies. To date, whole genome or whole exome sequencing has been performed in only 45 published DCIS cases in total, all using DNA from fresh frozen tissue.[4,14,15] Therefore, the cases used to date have been highly selected, derived from large mass forming tumors with adequate sufficient tissue to spare after diagnostic samples were taken.

To document the mutational profile of DCIS we performed massive parallel sequencing of a comprehensive panel of cancer-related genes on formalin-fixed, paraffin-embedded derived DNA from 20 DCIS tumors representing a full range of phenotypes. Our aims were to (1) document the mutational landscape of DCIS of different grades, hormone receptor and *ERBB2* status, (2) to identify differences between invasive and pre-invasive disease, (3) assess whether particular genomic alterations correlate with clinicopathological parameters, and (4) determine whether genomic profiling of DCIS can be implemented in the routine diagnostic setting.

Materials and methods

Clinicopathological Parameters of Patients and Tumors

Twenty DCIS cases were selected to represent a range of tumor characteristics (Table 1) and to have sufficient DNA quantity and quality to ensure successful sequencing. The median age of patients at surgery was 56.5 years (range 29–90 years, mean 56.5 years). Ten cases were of high nuclear grade, seven of intermediate nuclear grade, and three of low nuclear grade. Sixteen cases were estrogen receptor (ER)-positive, 14 progesterone receptor (PR)-positive, and five cases were *HER2* amplified. No long term clinical outcome data was available for these cases. Five cases had matched normal DNA available derived from adjacent breast tissue. Additional cases for *GATA3* sequencing were obtained from Royal Melbourne Hospital as previously described.[16] Approval for the study was obtained from the ethics committee of Peter MacCallum Cancer Centre (project numbers 02/26, 10/16, and 00/81).

Sample Processing and Sequencing

Areas of DCIS were microdissected from formalin-fixed paraffin embedded sections and DNA extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. A total of 500 ng per sample was used for sequencing. The KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, MA, USA) was used for library preparation with Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) for library clean-up. Target capture was performed using the SureSelectXT Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA), targeting the exons of 107 cancer-related genes (Supplementary Methods), including 61 genes of specific relevance to breast cancer, covering a total of 360 kb. Captured libraries were sequenced using a NextSeq500 sequencer (Illumina, San Diego, CA, USA).

Identification and Filtering of Variants

Sequence alignment and variant calling was performed by aligning the sequencer output to the reference genome using Burrows–Wheeler Aligner software[17] and variants were called using the SNPPET algorithm provided in the Agilent SureCall software package (Agilent Technologies). The variants identified by the Agilent SureCall algorithm were filtered for non-synonymous variants in exonic or essential splice site locations with a variant allele frequency of at least 10%. The allele frequency cut-off of 10% was chosen to minimize inclusion of sequencing artefacts related to formalin fixation.[18] As no matched normal DNA was submitted for sequencing, the 1000 Genomes database was used to exclude potential germline variants, which may represent non-pathogenic genetic variation in the population.[19] The sequencing reads in regions containing variants were then visually inspected using Integrative Genomics Viewer to exclude potentially artefactual variants, such as those occurring in variant-rich regions or variants identified exclusively at read ends.

Validation of Variants by Sanger Sequencing

The variants identified by massively parallel sequencing were validated by Sanger sequencing. Normal DNA, where available ($n=5$), was also subjected to Sanger sequencing alongside matching DCIS samples. Additional samples were sequenced for *GATA3* exons 5 and 6 using different primers with greater exon coverage. Sanger sequencing primers were designed using the Primer 3 tool.[20,21] Target sequences were amplified using primers and conditions listed in Supplementary Methods. The BigDye Terminator system (Applied Biosystems) was used for sequencing on a 3730 DNA Analyzer (Applied Biosystems). The sequencer output was viewed using Sequencer 4.8 software (Gene Codes Corporation, Ann Arbor, MI, USA) or Geneious (Biomatters, Auckland, New Zealand).

p53 Immunohistochemistry

Immunohistochemistry for p53 was performed using a Ventana BenchMark Ultra (Roche Diagnostics,

Table 1 DCIS cohort

Case	*Age*	*Nuclear grade*	*ER status*	*PR status*	*ERBB2 genetic status*	*PIK3CA*	*TP53*	*GATA3*	*Number variants*	*FGA1 (%)*	*NTAI2*	*Other mutations*	*Normal DNA*	*CN3 profile*
P121	73	Intermediate	Positive	Positive	Negative	Mutant	WT	Mutant	3	5.82	2	CBFB	No	Simple
P122	42	High	Negative	Negative	Amplified	WT	WT	WT	3	13.19	6	ATM, FGFR4, CDKN2A	No	Simple, amps on 17
P124	49	Intermediate	Positive	Positive	Negative	WT	WT	WT	2	5.03	2	PIK3R1	No	Simple, 2 amps
P125	64	Low	Positive	Positive	Negative	Mutant	WT	WT	2	1.68	0		No	Simple
P13	43	Intermediate	Positive	Positive	Negative	Mutant	WT	WT	1	NA	NA		Yes	NA
P19	54	High	Positive	Positive	Gain	Mutant	Mutant	WT	2	25.13	13		No	Complex
P27	64	High	Negative	Negative	Amplified	WT	Mutant	WT	2	47.80	18	PTCH1	No	Complex, multi amps
P43	53	High	Negative	Negative	Amplified	Mutant	Mutant	WT	2	21.48	13		Yes	Complex, multi amps
P44	59	High	Positive	Positive	Gain	WT	Mutant	Mutant	3	22.83	15	RUNX1	Yes	Complex, multi amps
P45	60	Intermediate	Positive	Positive	Negative	WT	WT	Mutant	1	7.25	3		No	Simple
P53	90	High	Positive	Positive	Amplified	Mutant	WT	WT	2	23.41	11	ARID1A	No	Complex, amps on 17
P54	63	High	Positive	Negative	Amplified	Mutant	Mutant	WT	4	24.14	19	TBX3, NF2	No	Complex, multi amps
P56	29	High	Negative	Negative	Amplified	WT	Mutant	WT	1	29.98	22		Yes	Complex, multi amps
P64	44	High	Positive	Positive	Amplified	Mutant	WT	Mutant	3	4.38	2	NBN	No	Simple, multi amps 1,8,17
P66	67	Intermediate	Positive	Negative	Mutant	WT	WT	Mutant	2	38.86	13	FGFR1	No	Complex
P79	44	High	Positive	Positive	Negative	WT	WT	Mutant	2	12.59	10	TSC2	No	Complex, amps on 17
P81	48	Low	Positive	Positive	Negative	Mutant	WT	Mutant	2	7.72	4	TSC2	No	Simple
P85	74	Low	Positive	Positive	Negative	Mutant	WT	WT	3	8.61	4	SF3B1, STK11	No	Simple
P89	50	Intermediate	Positive	Positive	Negative	Mutant	WT	Mutant	3	2.41	3	RUNX1	Yes	Simple, 1 amp
P92	60	Intermediate	Positive	Positive	Negative	Mutant	WT	Mutant	3	NA	NA	RAD51D	No	NA

Abbreviations: CN, copy number; FGA, fraction of the genome altered by copy number; NTAI, number of telomeric allelic imbalances; WT, wild type.

USA) on 3 μm formalin-fixed paraffin embedded tissue sections of tissue microarrays containing 0.5–2 mm cores of DCIS with up to eight-fold redundancy as described previously.[22,23] Antigen retrieval was performed in a high pH Ultra cell conditioning solution (CC1, Roche Diagnostics) for 32 min at 100 °C. Sections were incubated with the p53 antibody (Novocastra Liquid Mouse Monoclonal DO-7, Leica Biosystems) at 1/100 for 24 min at 36 °C. The On-board detection system, OptiView Universal DAB Detection Kit (Roche Diagnostics) was used in accordance with the manufacturer's instructions.

p53 immunohistochemical nuclear reactivity was scored for intensity (0 = no reactivity, 1 = weak, 2 = moderate, 3 = strong reactivity) and percentage of tumor cells positive to the closest 5%. Absence of p53 nuclear reactivity or nuclear reactivity in ≥ 60% of tumor cells was considered to be an abnormal pattern and suggestive of *TP53* mutation.[24] For cases represented by more than one tumor core, absence of p53 nuclear reactivity in all cores or ≥ 60% nuclear reactivity in any core was considered an abnormal pattern.

Copy Number Analysis

Copy number was generated using the CopywriteR tool in R[25] using a normal DNA sample that was included in the same sequencing batch as the baseline control. Copy number data from invasive breast carcinomas was downloaded as Level 3 data from the TCGA Data portal. Data were imported to Nexus (BioDiscovery) with thresholds of gain and loss of ± 0.2 $\log_2$ ratio. Fraction of the genome altered was calculated as the percentage of the genome in base pairs affected by gain or loss. Mutation data from TCGA was obtained from the cBio Portal.[26] The results published here are, in part, based upon data generated by TCGA project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov.

Results

Sequence Variants Detected

The median number of reads per sample was 20,054,319 (range 17 504 368–24 942 911; mean 20 300 466), with a median 16.0% being duplicate reads (range 10.2–27.7%; mean 16.6%). Total percentage mapped reads on average was 99.7% (median 99.7%; range 99.6–99.8%; mean 99.7%), and 60.1% were on-target reads (median 60.1%; range 51.8–66.1%; mean 60.1%). The median read-depth over the target regions was 1692 reads (range 1325–2031; mean 1651) and the average percentage of target bases with over 1000 reads was 99.1%. These parameters indicate very good technical

performance of the assay, which is expected to yield good quality and reliable data.

There were a total of 11 145 variants in 103 genes identified by the Agilent SureCall algorithm. The median number of variants per sample was 600.5 (range 150–652; mean 557.3 variants per sample). Filtering based on inspection of the sequencing reads and the 1000 Genomes data set resulted in a reduction in the number of variants (81.9% and 73.8% of variants, respectively), suggesting that most of the variants in the target regions were likely sequencing artefacts or single nucleotide polymorphisms. After filters were applied, 52 candidate somatic variants in 25 genes remained and were validated by Sanger sequencing (Table 2).

The majority of variants identified were validated by Sanger sequencing (48/50, 96%), apart from those occurring in *FGFR1* and *PALB2* (one variant each). It is unclear why these failed to validate, although with allele frequencies of 0.21 and 0.12, respectively, it is possible these were below the sensitivity of Sanger sequencing. The *GATA3* variants of two samples (chr10:8111479insC, allele frequency 0.26 and chr10:8115709_8115711delGAC, allele frequency 0.35) were not validated due to insufficient DNA. However, variants occurring in the same regions as these two *GATA3* variants were validated in other samples, giving confidence to the sequencing results (Table 2). Four variants in two patients were present in both the DCIS and matching normal DNA samples (Table 2). It is possible that other very rare germline variants remain in the validated variants, as some samples did not have matching normal DNA for comparison.

Mutated Genes in DCIS

Following Sanger validation there were a total of 46 variants in 19 genes over 20 samples. The cohort had a median of two variants per sample (range 1–4 variants per sample; mean 2.3 variants per sample). Forty variants (87%) were located in exons and the remaining six (13%) at essential splice sites. Similar to previous studies of DCIS,[14,15] single-nucleotide substitutions were the most frequent mutation type detected (33/46, 71.7%), with the most frequent substitution being C>T:G>A (14/33, 42.4%) (Figure 1).

PIK3CA was the most frequently mutated gene, harboring 12 mutations in 11 cases (11/20, 55%). All the *PIK3CA* mutations were missense mutations. There were ten known activating *PIK3CA* mutations, five in the helical domain (one E542K, one E545K, and three E545A mutations) and five located in the kinase domain (one H1047Y and four H1047R mutations). In addition there was a E542V mutation (helical domain), predicted to be damaging and deleterious by the SIFT[27] and PROVEAN[28] algorithms, respectively. One mutation (E726K) was predicted to be non-deleterious.[27,28] One case had co-existing E542K and E545A mutations, which were mutually exclusive on the sequencing reads with differing allele frequencies (0.23 and 0.10, respectively), suggesting that these mutations were present in separate clones rather than being mutations on separate alleles of the same clone.

The next most commonly mutated genes were *GATA3* (9/20, 45%) and *TP53* (6/20, 30%). In contrast to *PIK3CA*, only one of the nine *GATA3* mutations was a missense mutation, the remainder being four splice site mutations, three frameshift insertions, and one in-frame deletion. Both the missense mutation and in-frame deletion were predicted to be damaging and deleterious by the SIFT[27] and PROVEAN[28] algorithms, respectively. All but one of the mutations affected exons 5 and 6, consistent with the mutation types previously reported for this gene.[5] Of the *TP53* mutations, three were missense mutations, two splice site mutations and one was a frameshift deletion. The missense and splice site mutations are known deleterious mutations recorded in the IARC database.[29] Co-existing mutations of *GATA3* and *TP53* were present in one case, of *PIK3CA* and *TP53* in three cases, and of *PIK3CA* and *GATA3* in four cases (Figure 2). *RUNXI* (Figure 3) and *TSC2* were mutated in two cases each (2/20, 10%). The remaining genes were mutated in one case each (1/20, 5%, Table 2 and Figure 2). Of note, one case (P124) had two truncating mutations in *PIK3R1* suggestive of bi-allelic inactivation.

The proportion of *GATA3* mutated cases observed (45%) was considerably higher than that observed in studies of invasive breast cancer (4–22%).[4,5,30,31] We therefore evaluated *GATA3* mutation status in a further 91 pure DCIS cases using Sanger sequencing of exons 5 and 6 (as DNA was limited). Non-synonymous or splice region mutations were identified in 21/91 cases (23%, Supplementary Table 1), for a final frequency of 30/111 cases (27%). This frequency is significantly higher than observed in invasive breast cancer (TCGA 54/507, $P<0.0001$, Fisher's exact test), and remains significantly more when only those with known ER positive status are considered (25/70 (36%) compared with 54/390 (14%) in TCGA, $P<0.0001$). We observed a higher proportion of missense mutations in our cohort than expected compared with invasive cohorts (8/30), nevertheless it is possible these are rare polymorphisms not represented in any existing databases, as no matching normal DNA was available to remove germline variants. However, even considering only overtly deleterious mutations (splice site and truncating), we still observed a significantly higher frequency than expected from TCGA data (19% all, $P=0.02$, 27% ER-positive only, $P=0.007$).

We hypothesized that *GATA3* mutations might predict a better outcome, however, within ER positive cases treated by wide local excision, there was no association with recurrence ($P=0.20$, Cox log-rank test), although the power in this analysis

Table 2 Variants (filtered) detected by sequencing panel

Gene	*Genomic alteration (GRCg37/hg19)*	*Transcript*	*Transcript alteration*	*Protein alteration*	*Case*	*Validation*	*Germline result*
ARID1A	chr1:27100375C>T	NM_139135	c.4087C>T	p.Q1363X	P53	Validated	No matched normal
ATM	chr11:108196837G>A	NM_000051	c.6860G>A	p.G2287E	P122	Validated	No matched normal
BRCA2	chr13:32899219A>G	NM_000059	c.323A>G	p.N108S	P43	Validated	Present in normal
CBFB	chr16:67100619T>G	NM_001755	c.317T>G	p.V106G	P121	Validated	No matched normal
CDH1	chr16:68844136G>A	NM_004360	c.724G>A	p.V242I	P43	Validated	Present in normal
CDKN2A	chr9:21994285C>A	NM_058195	c.G46T	p.G16C	P122	Validated	No matched normal
ERBB2	chr17:37882896G>A	NM_004448	c.2954G>A	p.R985H	P66	Validated	No matched normal
ERCC2	chr19:45872243G>A	NM_000400	c.191C>T	p.P64L	P44	Validated	Present in normal
FGFR1	chr8:38287292T>C	NM_015850	c.266A>G	p.Q89R	P66	Not detected	No matched normal
FGFR3	chr4:1801503G>A	NM_000142	c.409G>A	p.G137R	P43	Validated	Present in normal
FGFR4	chr5:176517459C>T	XM_005265838	c.160C>T	p.R54C	P122	Validated	No matched normal
GATA3	chr10:8106075insC	NM_001002295	c.898dupC	p.L301Pfs*3	P64	Validated	No matched normal
GATA3	chr10:8111433-8111434delCA	NM_001002295	c.925-3_925-2delCA	splice site	P121	Validated	No matched normal
GATA3	chr10:8111433-8111434delCA	NM_001002295	c.925-3_925-2delCA	splice site	P44	Validated	Absent in normal
GATA3	chr10:8111433-8111434delCA	NM_001002295	c.925-3_925-2delCA	splice site	P81	Validated	No matched normal
GATA3	chr10:8111479insC	NM_001002295	c.968dupC	p.T324Hfs*29	P66	Insufficient DNA	No matched normal
GATA3	chr10:8115700A>C	NM_001002295	c.1051-2G>A	splice site	P89	Validated	Absent in normal
GATA3	chr10:8115709-8115711delGAC	NM_001002295	c.1058_1060delGAC	p.R353_P354delinsT	P92	Insufficient DNA	No matched normal
GATA3	chr10:8115712C>A	NM_001002295	c.1061C>A	p.P354H	P45	Validated	No matched normal
GATA3	chr10:8115950insACACCACCCCT	NM_001002295	c.1299_1300insACACCACCCCT	p.H434Tfs*46	P79	Validated	No matched normal
NBN	chr8:90967716G>A	NM_002485	c.1192C>T	p.Q398X	P64	Validated	No matched normal
NF2	chr22:30061039C>T	NM_000268	c.871C>T	p.R291C	P54	Validated	No matched normal
PALB2	chr16:23646349T>A	NM_024675	c.1518A>T	p.Q506H	P56	Not detected	No matched normal
PIK3CA	chr3:178936082G>A	NM_006218	c.1624G>A	p.E542K	P125	Validated	No matched normal
PIK3CA	chr3:178936083A>T	NM_006218	c.1625A>T	p.E542V	P53	Validated	No matched normal
PIK3CA	chr3:178936091G>A	NM_006218	c.1633G>A	p.E545K	P13	Validated	Absent in normal
PIK3CA	chr3:178936092A>C	NM_006218	c.1634A>C	p.E545A	P125	Validated	No matched normal
PIK3CA	chr3:178936092A>C	NM_006218	c.1634A>C	p.E545A	P89	Validated	Absent in normal
PIK3CA	chr3:178936092A>C	NM_006218	c.1634A>C	p.E545A	P92	Validated	No matched normal
PIK3CA	chr3:178938934G>A	NM_006218	c.2176G>A	p.E726K	P54	Validated	No matched normal
PIK3CA	chr3:178952084C>T	NM_006218	c.3139C>T	p.H1047Y	P64	Validated	No matched normal
PIK3CA	chr3:178952085A>G	NM_006218	c.3140A>G	p.H1047R	P121	Validated	No matched normal
PIK3CA	chr3:178952085A>G	NM_006218	c.3140A>G	p.H1047R	P19	Validated	No matched normal
PIK3CA	chr3:178952085A>G	NM_006218	c.3140A>G	p.H1047R	P43	Validated	Absent in normal
PIK3CA	chr3:178952085A>G	NM_006218	c.3140A>G	p.H1047R	P85	Validated	No matched normal
PIK3R1	chr5:67576453-67576453delA	NM_181523	c.732delA	p.K245Nfs*15	P124	Validated	No matched normal
PIK3R1	chr5:67591114-67591117delCCTT	NM_181504	c.897_900del	p.D299Efs*4	P124	Validated	No matched normal
PTCH1	chr9:98241349G>A	NM_000264	c.1148C>T	p.S383L	P27	Validated	No matched normal
RAD51D	chr17:33446607-33446608delCA	NM_002878	c.25_26del	p.C9Pfs*61	P92	Validated	No matched normal
RUNX1	chr21:36206724insG	NM_001001890	c.706dupC	p.Q237Sfs*336	P89	Validated	Absent in normal
RUNX1	chr21:36252866G>C	NM_001001890	c.415C>G	p.R139G	P44	Validated	Absent in normal
SF3B1	chr2:198265515G>C	NM_012433	c.2642C>G	p.A881G	P85	Validated	No matched normal
STK11	chr19:1206917A>C	NM_000455	c.5A>C	p.E2A	P85	Validated	No matched normal
TBX3	chr12:115118738insT	NM_005996	c.602_603insA	p.V202Rfs*38	P54	Validated	No matched normal
TP53	chr17:7577024-7577048delTTAGTGC TCCCTGGGGGCAGCTCGT	NM_000546.5	c.890_914delACGAGCTGCCC CCAGGGAGCACTAA	p.H297Rfs*40	P56	Validated	Absent in normal
TP53	chr17:7577105G>A	NM_000546.5	c.833C>T	p.P278L	P19	Validated	No matched normal
TP53	chr17:7577131G>C	NM_000546.5	c.807C>G	p.S269R	P43	Validated	Absent in normal
TP53	chr17:7577156C>A	NM_000546.5	c.783-1G>T	splice site	P27	Validated	No matched normal
TP53	chr17:7577610T>G	NM_000546.5	c.673-2A>C	splice site	P44	Validated	Absent in normal
TP53	chr17:7578496A>G	NM_000546.5	c.434T>C	p.L145P	P54	Validated	No matched normal
TSC2	chr16:2134485G>A	NM_000548	c.4262G>A	p.G1421E	P79	Validated	No matched normal
TSC2	chr16:2134623C>T	NM_000548	c.4400C>T	p.A1467V	P81	Validated	No matched normal

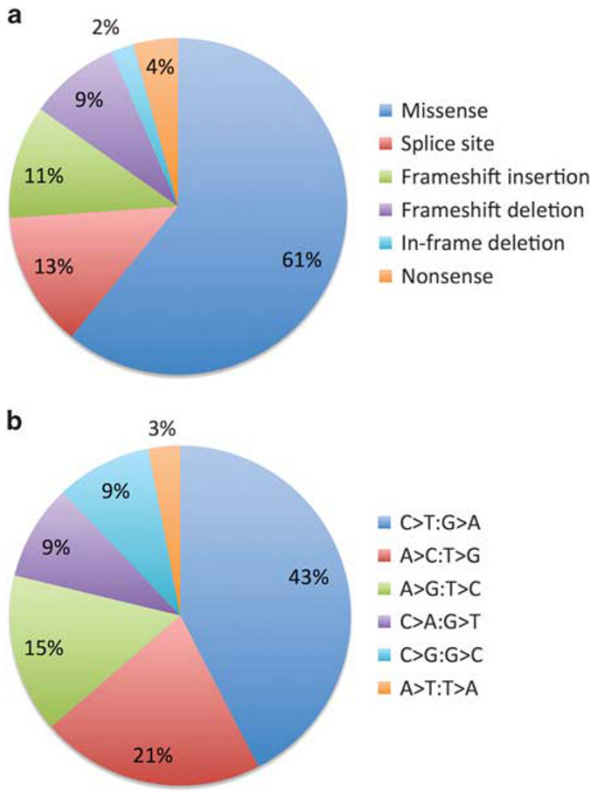

Figure 1 Mutation types in DCIS. (**a**) Distribution of mutation types. (**b**) Distribution of single nucleotide substitutions. Both consider only filtered variants; known germline variants excluded.

was limited, with only seven events and no cases of invasive recurrence.

Mutations and DCIS Phenotype

There was no difference in the median number of mutations or the median number of genes mutated by nuclear grade, ER status, PR status, or *HER2* amplification status (Table 3). However, mutations of some genes did show an association with DCIS phenotype. *TP53* mutations occurred exclusively in high nuclear grade DCIS ($P=0.011$) and occurred more frequently in PR-negative DCIS (4/6, 66.7%) compared with PR-positive DCIS (2/14, 14.3%) ($P=0.037$). There was a trend of *TP53* mutations occurring more frequently in ER-negative DCIS cases but this association did not reach statistical significance (3/4, 75% ER-negative *vs* 3/16, 18.8% ER-positive DCIS, $P=0.061$). No association was observed between the presence of *PIK3CA* mutations and DCIS phenotype. Mutations of other genes occurred in too few cases for a meaningful assessment of association with phenotypic features.

Copy Number Alterations

We undertook genome-wide copy number analysis using the CopywriteR algorithm, which uses the off-target reads to estimate copy number.[25] Two cases, P13 and P92, gave poor quality copy number output and were excluded from further analysis. All the remaining 18 cases had at least one copy number event, and the copy number profiles generated were similar to those we previously reported for pure DCIS[32], including common gains on 1q, 8q, 17q, and 20q and frequent losses on 8p, 11q, 16q, and 17p. There was good correlation between the copy number values generated by the on- and off-target reads for the genes in the panel (Spearman $r=0.96$, $P<0.0001$, Figure 4), and *ERBB2* amplifications detected by sequencing were consistent with SISH data.

Copy number profiles were evaluated for various measures of genomic instability, including overall aberration type, the fraction of the genome altered by copy number and a measure of homologous recombination deficiency, the number of telomeric allelic imbalances.[33] Seven cases had simple profiles (all were low or intermediate grade and ER positive), three cases had a background of simple copy number change but with two or more high-level amplifications, and the remaining eight cases had highly complex profiles, most with high level amplifications. No cases had chromothripsis. The fraction of the genome altered by copy number ranged from 2 to 47% (median 12.9%). *TP53* mutant tumors had a significantly higher fraction of the genome altered by copy number than wild-type tumors (Figure 4, $P=0.005$, Mann–Whitney two-tail test). The median telomeric imbalance score for the 18 DCIS cases was 8 (range 0–22), slightly lower than the median of 12 in invasive breast cancer.[34] In our previous cohort of 53 DCIS with molecular inversion probe array data, the median was 8.5, thus we do not think the score is reduced by not having allelic imbalance information available. The telomeric imbalance score was significantly associated with *TP53* mutation (Figure 4, $P=0.001$) and high nuclear grade ($P=0.01$).

We integrated our copy number and mutation data to investigate associations of particular mutations with copy number profiles. All *TP53* mutant tumors had a complex copy number profile and were significantly enriched for gain of *ERBB2*, 3q, and 20q and loss of 9p, 17p, and X (Figure 4). In invasive breast cancers from TCGA, 3q gain was also strongly associated with *TP53* mutation but the other regions were not significant at a threshold of >25% difference in frequency. All *TP53* mutant tumors had either amplification ($n=4$) or gain ($n=2$) of *ERBB2*, similar to Abba *et al*[14] (amplified $n=4$, gain $n=1$). To further investigate this association, we compared p53 immunohistochemistry with *ERBB2* amplification status by silver *in situ* hybridization in a wider DCIS cohort, using tissue microarrays including both pure DCIS ($n=187$) and DCIS associated with invasive breast cancer ($n=19$). Abnormal p53 staining was observed in 68/187 (36.0%) pure DCIS and 9/19 (47%) mixed DCIS, with the majority of abnormal cases (57%) showing

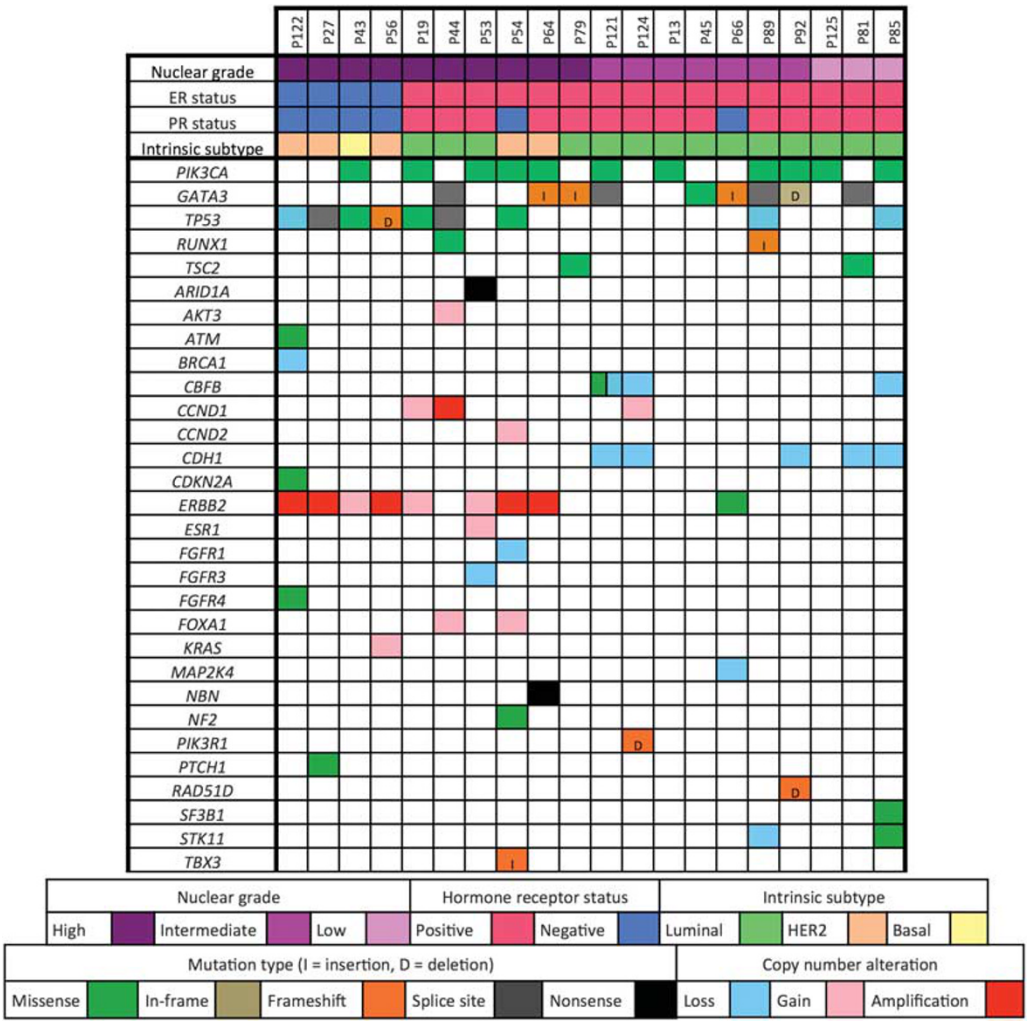

Figure 2 Mutations and copy number alterations identified in DCIS (germline variants excluded).

over-expression (Table 4). There was no significant difference in p53 positivity between pure and mixed DCIS ($P=0.46$). A significant association of p53 abnormal staining with *ERBB2* amplification was observed in pure DCIS ($P=0.033$). Half of *ERBB2* amplified pure DCIS cases were p53 abnormal, compared with 30% of *ERBB2* non-amplified. Considering only p53 over-expression, 38% of *ERBB2* amplified were p53 over-expressing, compared to 16% of *ERBB2* non-amplified ($P=0.003$). Abnormal p53 protein was also significantly associated with high nuclear grade, ER negativity, and PR negativity (Table 4).

Twelve of the twenty DCIS cases submitted for panel sequencing had p53 immunohistochemistry data. All five *TP53* mutants with protein data were classified as p53 abnormal (either entirely negative for a truncating mutation or strongly positive in >60% of cells for a missense mutation; sensitivity 100%). Six of the seven cases classified as *TP53* wildtype by sequencing were p53 normal by immunohistochemistry (specificity 85.7%), with the one discordant case negative for protein. Immunohistochemistry is therefore a fair proxy for mutation status in DCIS using these thresholds, given likely tumor heterogeneity.

PIK3CA-mutated tumors were more likely to have loss of chromosome 18, while *GATA3*-mutated tumors were less likely to have loss of 8p. Neither observation was significant in TCGA invasive breast cancer data. Neither *PIK3CA* nor *GATA3* mutations were associated with a significant difference in the fraction of the genome altered by copy number or the number of telomeric imbalances.

Discussion

Little is known about the mutations that occur in DCIS. We therefore used massively parallel sequencing of a targeted panel of 107 cancer-related genes to determine mutations and copy number alterations in 20 DCIS cases of various phenotypes. The panel was selected to represent frequently mutated breast

cancer drivers and clinically actionable genes. The technique was shown to be robust for the detection of genomic alterations in formalin-fixed paraffin embedded derived DNA (96.0% validation rate). Previous studies examining mutations of a broad panel of genes in DCIS have utilized DNA from

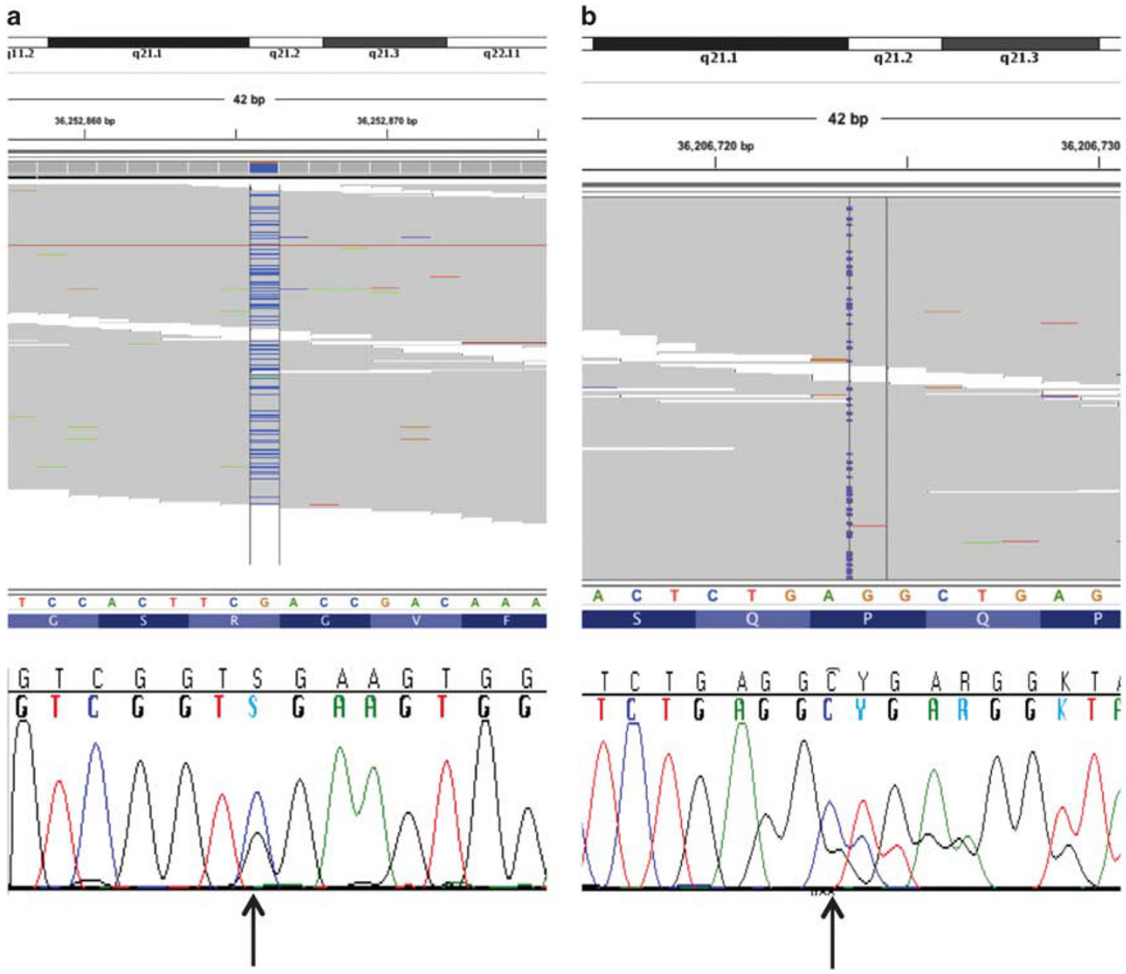

Figure 3 *RUNX1* mutations. (**a**) Sample P44 with chr21:36252866G>C. (**b**) Sample P89 with chr21:36206724 insG. In both panels the Integrated Genome Viewer plot of the targeted sequencing bam file is shown at top, while the Sanger sequencing validation electrophoretogram is shown below.

Table 3 Associations between mutations and DCIS phenotype

	High grade	*Non-high grade*	P-*value*	*ER+*	*ER-*	P-*value*	*PR+*	*PR-*	P-*value*	*HER2 amplified*	*HER2 non-amplified*	P-*value*
Median number of mutations per case	2	2	0.971	2	2	0.554	2	2	0.968	2	2	0.800
Median number of mutated genes per case	2	2	0.579	2	2	0.750	2	2	0.718	2	2	0.612
GATA3												
Mutation present	3	6	0.370	9	0	0.094	8	1	0.157	1	8	0.319
No mutation	7	4		7	4		6	5		4	7	
PIK3CA												
Mutation present	5	6	1.000	10	1	0.285	9	2	0.336	2	9	0.617
No mutation	5	4		6	3		5	4		3	6	
TP53												
Mutation present	6	0	**0.011**	3	3	0.061	2	4	**0.037**	3	3	0.131
No mutation	4	10		13	1		12	2		2	12	

Bold values indicate statistical significance at $P < 0.05$.

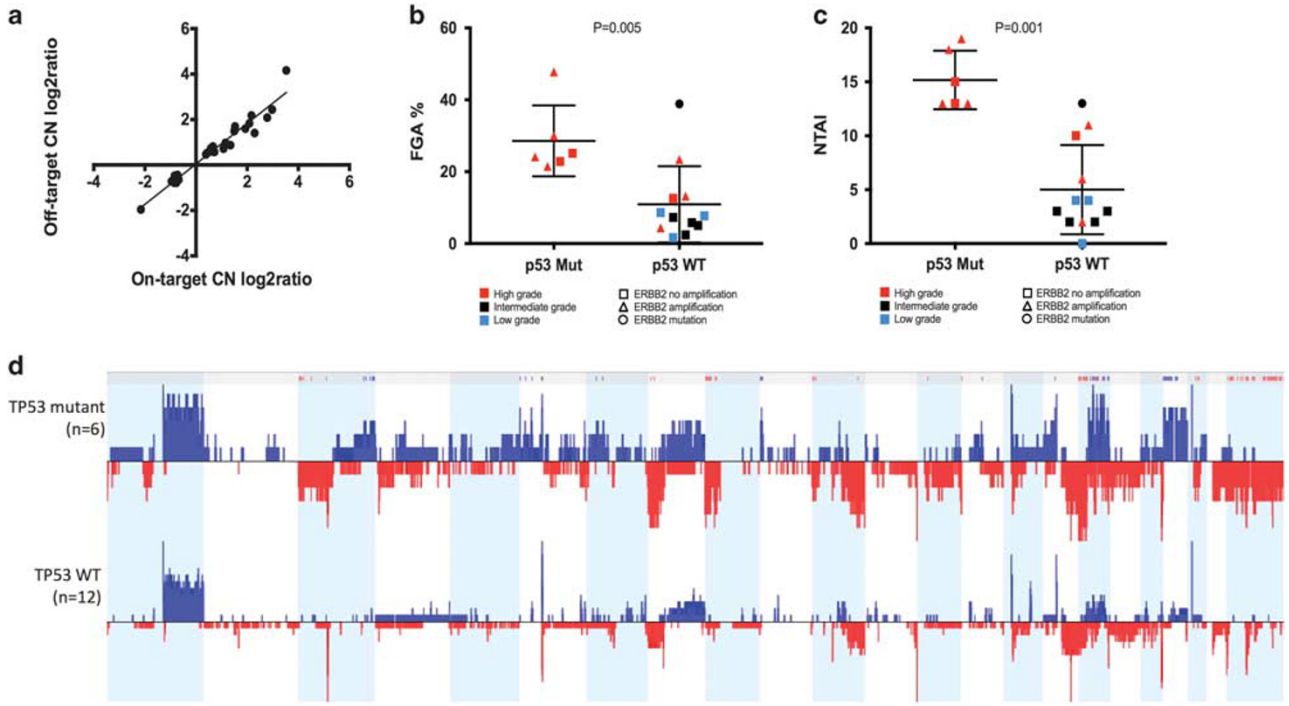

Figure 4 Integration of mutation with copy number data. (**a**) Good correlation between on- and off-target copy number predictions was observed. (**b**) Fraction of the genome altered by copy number (FGA) by *TP53* mutation status. (**c**) Number of telomeric imbalances (NTAI) by *TP53* mutation status. (**d**) Comparison of copy number profiles of *TP53* mutant *vs* wild-type tumors (blue = gain, red = loss). Bar at the top indicates statistical significance at $P < 0.05$.

fresh-frozen tumor specimens, but this study shows that interrogation of formalin-fixed paraffin embedded-derived DNA can now be applied to a variety of DCIS cases from hospital-based series, including cases archived for up to nine years at the time of DNA extraction.

The rate of variants in DCIS samples was higher (5.6 mutations/Mb) than that previously reported for DCIS (1.61 mutations/Mb)[14] and invasive breast cancer (1.66 mutations/Mb)[4] and is likely to be due to the use of a panel focussed on genes known to be mutated in breast cancer. Although potential germline variants might have contributed to this (these were unable to be completely excluded as few cases had matched normal DNA available), the number of these germline variants is expected to be small, as stringent population filters based on large public databases were applied.

The most prevalent single-base substitution observed was the C>T:G>A alteration, accounting for 43% of single-nucleotide substitutions detected. The C>T:G>A alteration forms a major part of three of the five mutational signatures identified in breast cancer[35] and is thought to occur through the spontaneous deamination of 5-methyl-cytosine which accumulates with age[35,36] and overactivity of the APOBEC family of cytidine deaminases.[35,37] Although C>T:G>A alterations are also characteristic of formalin-related DNA damage[18] the proportion of C>T:G>A alterations is similar to that reported in fresh-frozen samples of invasive breast cancer (~40%)[4] and DCIS (~50%)[14,15], so the C>T:G>A alterations detected are unlikely to represent formalin-fixation artefacts.

Mutations of 19 genes were identified including those involved in DNA repair and cell cycle control (*ARID1A, ATM, CDKN2A, NBN, RAD51D, TP53*), the PI3K/AKT/mTOR pathway (*PIK3CA, PIK3R1, TSC2*), the Hedgehog pathway (*PTCH1*), transcription factors (*CBFB, GATA3, RUNX1, TBX3*), receptor tyrosine kinases (*ERBB2, FGFR4*), serine/threonine-protein kinases (*STK11*), other tumor suppressor genes (*NF2*), and splicing factors (*SF3B1*), suggesting that multiple mechanisms are disrupted in the development of DCIS. Notably, unlike Abba *et al*,[14] we only saw one case that had no common driver (ie, wild-type for *TP53*, *PIK3CA*, *ERBB2,* and *GATA3*). Although this case (P124) had a relatively stable genome (5% affected by copy number), the copy number changes included a *CCND1* amplification, *CBFB* loss, and *CDH1* loss. The sample also carried two *PIK3R1* truncating mutations. Thus, DCIS cases of all phenotypes carry mutational drivers reminiscent of invasive breast cancer, although the frequency of some events, such as *GATA3* mutations, may vary. Analysis of more DCIS cases with matching normal DNA will be required to determine whether the more rarely mutated genes are indeed more frequently affected in DCIS than invasive breast cancer.

The prevalence of *PIK3CA* mutations in this cohort (55%) is somewhat higher than that previously

Table 4 p53 immunohistochemistry

	Pure DCIS (n = *187*)			*Mixed DCIS* (n = *19*)		
	p53 abnormal	*p53 normal*	P-*value* (*Fisher's exact test*)	*p53 abnormal*	*p53 normal*	P-*value* (*Fisher's exact test*)
Nuclear grade						
High grade	43	55	< 0.001	8	3	0.004
Non-high grade	13	56		0	7	
ER status						
ER positive	30	90	< 0.0001	3	8	0.11
ER negative	32	24		4	1	
PR status						
PR positive	17	74	< 0.0001	0	5	0.044
PR negative	43	43		6	4	
HER2 status						
Amplified	19	19	0.033	3	2	0.62
Non-amplified	42	99		5	7	
Clinical outcome						
Recurred	3	9	0.65	–	–	
No recurrence	4	6				

reported for DCIS (17–48%)[7,8,14,15,38,39] and invasive breast cancers (25–36%).[4,5,40,41] The relatively small number of cases included in this and other DCIS each studies (involving between six and 202 cases),[7,8,14,15,38,39] may account for the variation in prevalence. Alternatively, the high read-depth of our targeted sequencing may have led to increased detection rates compared with exome or Sanger studies: four DCIS cases had a *PIK3CA* variant frequency of < 20%, which could have been missed by less sensitive methods. The role of *PIK3CA* mutations as markers of tumor progression in DCIS is uncertain. *PIK3CA* mutations in invasive breast cancers have been associated with favorable tumor features[41] and better prognosis in patients with ER-positive, HER2-negative tumors;[40] however, we saw no association of *PIK3CA* mutations with clinicopathological features in our cohort.

The *GATA3* mutations were mostly observed in ER-positive cases (93%), in keeping with the reported association of *GATA3* mutations with luminal breast cancer.[5,42] Only one case had both *GATA3* and *TP53* mutations, reflecting a pattern reported in invasive breast tumors where mutations of these two genes rarely co-exist.[42] Four DCIS cases had both *GATA3* and *PIK3CA* mutations, in contrast to invasive carcinomas in which mutations of these two genes are almost mutually exclusive.[1,42] In one case the allele frequencies suggest that the *GATA3* and *PIK3CA* mutations could be present in different clones.

The prevalence of *GATA3* mutations in the current DCIS cohort is significantly higher than that reported in invasive breast cancer (4–22%)[4,5,30,31] and in DCIS. Recently, Abba *et al*[14] reported a *GATA3* mutation rate of 7% in 30 pure high grade DCIS cases while Kim *et al*[15] observed *GATA3* mutation in one low grade DCIS out of a cohort of in six DCIS lesions of various nuclear grades (16.7%). If the prevalence of *GATA3* mutations observed in the current cohort is truly reflective of the general DCIS population, the high prevalence of *GATA3* mutations in DCIS compared with invasive carcinomas suggests that *GATA3* mutations, like *HER2* amplification, are not selected for during the transition to invasive disease and perhaps may indicate better prognosis, similar to that observed in invasive carcinomas.[42] Our extended cohort did not have sufficient invasive recurrences to be able to test this hypothesis.

TP53 mutations were present in exclusively in high grade DCIS, and half the mutations were missense; both these features are consistent with previous reports of *TP53* mutations in both invasive and *in situ* breast cancer.[1,4–6,12,13] We also noted an association of abnormal p53 protein with *ERBB2* amplification. In invasive breast cancer from TCGA, *TP53* mutation is also significantly associated with *ERBB2* amplification ($P < 0.0001$), and the proportion of *ERBB2* amplified cases with *TP53* mutation (61/118, 52%) is similar to our pure DCIS cohort (50%), despite the different methods used (exome *vs* immunohistochemistry and SNP arrays *vs in situ* hybridization for p53 and *ERBB2*, respectively). A key question in DCIS biology is the relatively higher proportion of cases with *ERBB2* amplification compared to invasive cases: our data would not support a role for *TP53* in explaining this difference.

We identified *RUNX1* mutations in two DCIS cases, which has not previously been reported in DCIS but occurs in 4–5% of invasive breast carcinomas,[4,5] where mutations are associated with luminal B expression profiles and high grade tumors[1]. The DCIS cases with *RUNX1* mutations were both ER positive, HER2 negative and high or intermediate grade, consistent with a luminal B phenotype. *RUNX1* is required for ER localization and regulation of target genes[43] and is believed to function as a tumor suppressor gene in breast cancer.[44] We also observed a mutation of *CBFB* which was mutually exclusive with *RUNX1* mutation, consistent with their role as subunits of a heterodimeric transcription factor. The *CBFB*-mutated case was also ER positive, HER2 negative and intermediate grade. The mutually exclusive relationship between mutations of *RUNX1* and *CBFB*, and also *PIK3R1* and *PIK3CA* is observed in invasive breast cancer.[4,5]

As the panel assays genes predominantly altered in invasive breast cancer rather than an agnostic exome or genome approach, mutations unique to DCIS were unable to be identified. Nevertheless, a targeted approach has several advantages, including high read-depth, which enabled detection of mutations even at allele frequencies of 10%. In addition, the cost was relatively low and bioinformatics analysis quick and straightforward. We were also able to obtain good quality copy number data from almost all samples, which enabled copy number drivers such as *CCND1* amplification and *MAP2K4* homozygous deletion to be observed. Copy number alterations of *MAP2K4* have not previously been reported in DCIS.

In conclusion, this study provides a snapshot of the mutational profiles of DCIS, incorporating both copy number and somatic point mutations. The entire spectrum of mutations in DCIS and even in an individual tumor is unlikely to have been documented in this study due to the small cohort, targeted gene panel approach and limited sampling of the lesions. Nonetheless, mutations were present in all samples suggesting that mutational processes have a role in DCIS biology. While generally similar mutational patterns to invasive breast carcinomas were observed, there was a surprisingly high prevalence of *GATA3* mutations. Given that *GATA3* mutations are associated with improved survival in patients with invasive breast cancer, it is hypothesized that *GATA3* mutations in DCIS may be a marker of less aggressive behavior. In contrast, *TP53* mutations were associated with adverse tumor characteristics. The role of copy number alteration, telomeric imbalance and specific gene mutations as prognostic biomarkers will require testing in a suitable cohort of DCIS with outcome data available. The panel testing approach used here will be an appropriate methodology to address these questions.

Acknowledgments

This study was supported by the Australian National Health and Medical Research Council (ID #50950, #566603, and APP1063092) and the National Breast Cancer Foundation (NBCF). TK was supported by a Melbourne University International Scholarship. We thank Dr Siobhan Hughes for technical assistance.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Ellis MJ, Ding L, Shen D, *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 2012;486:353–360.
2 Curtis C, Shah SP, Chin SF, *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–352.
3 Stephens PJ, McBride DJ, Lin ML, *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 2009;462:1005–1010.
4 Banerji S, Cibulskis K, Rangel-Escareno C, *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486:405–409.
5 Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
6 Mao X, Fan C, Wei J, *et al.* Genetic mutations and expression of p53 in non-invasive breast lesions. Mol Med Rep 2010;3:929–934.
7 Sakr RA, Weigelt B, Chandarlapaty S, *et al.* PI3K pathway activation in high-grade ductal carcinoma *in situ*–implications for progression to invasive breast carcinoma. Clin Cancer Res 2014;20:2326–2337.
8 Miron A, Varadi M, Carrasco D, *et al.* PIK3CA mutations in *in situ* and invasive breast carcinomas. Cancer Res 2010;70:5674–5678.
9 Dunlap J, Le C, Shukla A, *et al.* Phosphatidylinositol-3-kinase and AKT1 mutations occur early in breast carcinoma. Breast Cancer Res Treat 2010;120:409–418.
10 Munn KE, Walker RA, Menasce L, *et al.* Mutation of the TP53 gene and allelic imbalance at chromosome 17p13 in ductal carcinoma *in situ*. Br J Cancer 1996;74:1578–1585.
11 Chitemerere M, Andersen TI, Holm R, *et al.* TP53 alterations in atypical ductal hyperplasia and ductal carcinoma *in situ* of the breast. Breast Cancer Res Treat 1996;41:103–109.
12 Done SJ, Eskandarian S, Bull S, *et al.* p53 missense mutations in microdissected high-grade ductal carcinoma *in situ* of the breast. J Natl Cancer Inst 2001;93:700–704.
13 Zhou W, Muggerud AA, Vu P, *et al.* Full sequencing of TP53 identifies identical mutations within *in situ* and invasive components in breast cancer suggesting clonal evolution. Mol Oncol 2009;3:214–219.
14 Abba MC, Gong T, Lu Y, *et al.* A molecular portrait of high-grade ductal carcinoma *in situ*. Cancer Res 2015;75:3980–3990.
15 Kim SY, Jung SH, Kim MS, *et al.* Genomic differences between pure ductal carcinoma *in situ* and

synchronous ductal carcinoma *in situ* with invasive breast cancer. Oncotarget 2015;6:7597–7607.
16 Gorringe K, Hunter S, Pang J, *et al.* Copy number analysis of DCIS with and without recurrence. Mod Pathol 2015;28:1174–1184.
17 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754–1760.
18 Wong SQ, Li J, Tan AY, *et al.* Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics 2014;7:23.
19 Genomes Project C, Abecasis GR, Auton A, *et al.* An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
20 Untergasser A, Cutcutache I, Koressaar T, *et al.* Primer3–new capabilities and interfaces. Nucleic Acids Res 2012;40:e115.
21 Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics 2007;23:1289–1291.
22 Pang JM, Deb S, Takano EA, *et al.* Methylation profiling of ductal carcinoma *in situ* and its relationship to histopathological features. Breast Cancer Res 2014; 16:423.
23 Zardawi SJ, Zardawi I, McNeil CM, *et al.* High Notch1 protein expression is an early event in breast cancer development and is associated with the HER-2 molecular subtype. Histopathology 2010;56:286–296.
24 Yemelyanova A, Vang R, Kshirsagar M, *et al.* Immunohistochemical staining patterns of p53 can serve as a surrogate marker for TP53 mutations in ovarian carcinoma: an immunohistochemical and nucleotide sequencing analysis. Mod Pathol 2011;24:1248–1253.
25 Kuilman T, Velds A, Kemper K, *et al.* CopywriteR: DNA copy number detection from off-target sequence data. Genome Biol 2015;16:49.
26 Cerami E, Gao J, Dogrusoz U, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401–404.
27 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4: 1073–1081.
28 Choi Y, Sims GE, Murphy S, *et al.* Predicting the functional effect of amino acid substitutions and indels. PLoS ONE 2012;7:e46688.
29 Petitjean A, Mathe E, Kato S, *et al.* Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat 2007;28: 622–629.
30 Arnold JM, Choong DY, Thompson ER, *et al.* Frequent somatic mutations of GATA3 in non-BRCA1/BRCA2 familial breast tumors, but not in BRCA1-, BRCA2- or sporadic breast tumors. Breast Cancer Res Treat 2010;119:491–496.
31 Usary J, Llaca V, Karaca G, *et al.* Mutation of GATA3 in human breast tumors. Oncogene 2004;23:7669–7678.
32 Gorringe KL, Hunter SM, Pang JM, *et al.* Copy number analysis of ductal carcinoma *in situ* with and without recurrence. Mod Pathol 2015;28:1174–1184.
33 Birkbak NJ, Wang ZC, Kim JY, *et al.* Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer Discov 2012;2: 366–375.
34 Marquard AM, Eklund AC, Joshi T, *et al.* Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. Biomark Res 2015;3:9.
35 Alexandrov LB, Nik-Zainal S, Wedge DC, *et al.* Signatures of mutational processes in human cancer. Nature 2013;500:415–421.
36 Pfeifer GP. Mutagenesis at methylated CpG sequences. Curr Top Microbiol Immunol 2006;301:259–281.
37 Nik-Zainal S, Alexandrov LB, Wedge DC, *et al.* Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979–993.
38 Ang DC, Warrick AL, Shilling A, *et al.* Frequent phosphatidylinositol-3-kinase mutations in proliferative breast lesions. Mod Pathol 2014;27:740–750.
39 Li H, Zhu R, Wang L, *et al.* PIK3CA mutations mostly begin to develop in ductal carcinoma of the breast. Exp Mol Pathol 2010;88:150–155.
40 Loi S, Haibe-Kains B, Majjaj S, *et al.* PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor-positive breast cancer. Proc Natl Acad Sci USA 2010;107: 10208–10213.
41 Kalinsky K, Jacks LM, Heguy A, *et al.* PIK3CA mutation associates with improved outcome in breast cancer. Clin Cancer Res 2009;15:5049–5059.
42 Jiang YZ, Yu KD, Zuo WJ, *et al.* GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival. Cancer 2014;120:1329–1337.
43 Stender JD, Kim K, Charn TH, *et al.* Genome-wide analysis of estrogen receptor alpha DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation. Mol Cell Biol 2010;30:3943–3955.
44 Janes KA. RUNX1 and its understudied role in breast cancer. Cell Cycle 2011;10:3461–3465.

Supplementary Information accompanies the paper on Modern Pathology website (http://www.nature.com/modpathol)