18-602.R1

Poulsom Edited UK English

ORIGINAL PAPER

Phenotypic and molecular dissection of Metaplastic Breast Cancer and the prognostic implications.
[Change of title agreed with authors]

Amy Ellen McCart Reed¹, Emarene Kalaw¹, Katia Nones², Mark Bettington¹, Malcolm Lim¹, James Bennett^{1,3}, Kate Johnstone^{1,3}, Jamie Rose Kutasovic¹, Jodi Marie Saunus¹, Stephen Kazakoff², Qinying Xu², Scott Wood², Oliver Holmes², Conrad Leonard², Lynne Estelle Reid¹, Debra Black¹, Colleen Niland¹, Kaltin Ferguson¹, Irma Gresshoff¹, Ashwini Raghavendra¹, Kate Harvey⁴, Caroline Cooper⁴, Cheng Liu^{1,3}, Lauren Kalinowski^{1,3}, Andrew Scott Reid^{1,3}, Morgan Davidson^{1,3}, John V Pearson², Nirmala Pathmanathan⁵, Gary Tse⁶, David Papadimos⁷, Rajadurai Pathmanathan⁸, Gavin Harris⁹, Rin Yamaguchi¹⁰, Puay Hoon Tan¹¹, Stephen Fox¹², Sandra A O'Toole⁴, Peter Thomas Simpson¹, Nicola Waddell², Sunil R Lakhani^{1,3*}.

¹ UQ Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, Brisbane Australia.

² QIMR Berghofer Medical Research Institute, Herston, Brisbane, Australia.

³ Pathology Queensland, The Royal Brisbane & Women's Hospital, Herston, Brisbane Australia.

⁴ Garvan Institute of Medical Research and the Kinghorn Cancer Centre, 384 Victoria Street, Darlinghurst, NSW, 2010, Australia.

⁵ Westmead Breast Cancer Institute; University of Sydney.

⁶ Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, Hong Kong.

⁷ Department of Histopathology, Sullivan Nicolaides Pathology, Bowen Hills, Queensland, Australia.

⁸ Sime Darby Medical Centre, 47500 Selangor, Malaysia.

⁹ Canterbury Health Laboratories, Christchurch, New Zealand/Dept of Molecular Medicine and Pathology, University of Auckland, New Zealand.

¹⁰ Department of Pathology and Laboratory Medicine, Kurume University Medical Center, 155-1 Kokubu, Kurume-shi 839-0863, Japan.

¹¹ Division of Pathology, Singapore General Hospital, Singapore.

¹² Peter MacCallum Cancer Centre, Melbourne, Australia.

Short running title: Prognostic features of Metaplastic breast cancer.

The authors declare they have no conflicts of interest.

* Correspondence to: Prof Sunil Lakhani, The University of Queensland Centre for Clinical Research Level 6 Building 71/918, The Royal Brisbane & Women's Hospital, Herston, Brisbane QLD 4029, Australia. Phone: +61 7 3346 6052. E-mail: s.lakhani@uq.edu.au

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/path.5184](https://doi.org/10.1002/path.5184)

Abstract

Metaplastic Breast Carcinoma (MBC) is relatively rare but accounts for a significant proportion of global breast cancer mortality. This group is extremely heterogeneous and by definition exhibits metaplastic change to squamous and/or mesenchymal elements, including spindle, squamous, chondroid, osseous and rhabdomyoid features. Clinically, patients are more likely to present with large primary tumours (higher stage), distant metastases and overall, have shorter 5-year survival compared to Invasive Carcinomas of no special type. The current World Health Organisation (WHO) diagnostic classification for this cancer type is based purely on morphology – the biological basis and clinical relevance of its seven sub-categories are currently unclear. By establishing the Asia-Pacific MBC (AP-MBC) consortium, we amassed a large series of MBC (n=347) and analysed the mutation profile of a subset, expression of 14 breast cancer biomarkers and clinicopathologic correlates, contextualizing our findings within the WHO guidelines. The most significant indicators of poor prognosis were: large tumour size (T3; p=0.004), loss of cytokeratin expression (lack of staining with pan-Cytokeratin AE1/3 antibody; p=0.007), EGFR overexpression (p=0.01) and for ‘mixed’ MBC, the presence of more than three distinct morphological entities (p=0.007). Conversely, fewer morphological components and EGFR negativity were favourable indicators. Exome sequencing of 30 cases confirmed enrichment of *TP53* and *PTEN* mutations, and intriguingly, concurrent mutations of *TP53*, *PTEN* and *PIK3CA*. Mutations in neurofibromatosis-1 (*NF1*) were also over-represented (16.7% MBC compared to ~5% of breast cancers overall; enrichment p=0.028; mutation significance p=0.006 (OncodriveFM)), consistent with published case reports implicating germline *NF1* mutations in MBC risk. Taken together, we propose a

practically minor but clinically significant modification to the guidelines: all WHO_1 mixed type tumours should have the number of morphologies present recorded, as a mechanism for refining prognosis, and that EGFR and pan-cytokeratin expression are important prognostic markers.

Keywords

Metaplastic, Breast cancer, Heterogeneity, Squamous, Spindle, Genomics, Epithelial to mesenchymal transition

Introduction

Metaplastic breast carcinoma (MBC) accounts for 0.2–5% of invasive breast cancers. While rare in terms of population incidence, they usually carry a poor prognosis and thus contribute significantly to global breast cancer mortality. Histologically, MBC is defined by the presence of metaplastic change to squamous and/or mesenchymal elements, but as a group and individually, they are highly heterogeneous, exhibiting spindle, squamous, chondroid, osseous and/or rhabdomyoid elements [1]. MBCs are usually triple-negative (TN; lacking expression of oestrogen receptor (ER), progesterone receptor (PR) and HER2), but carry a worse prognosis than other grade-matched TN breast cancers, with twice the risk of local recurrence, shorter disease-free and overall survival (DFS/OS) [2, 3]. Consistent with their TN phenotype, they are less likely than IC-NST (Invasive Carcinoma - no special type) to present with lymph node involvement, yet more likely to present with distant metastases (reviewed in [4]).

Being uncommon in individual institutions around the world has limited opportunities to characterise the MBC molecular genetic landscape in detail, but details are beginning to emerge from small- and moderately-sized case series. For example, the majority over-express EGFR, underpinned by gene amplification in ~30% of cases [5]. Consistent with mesenchymal morphology, the MBC transcriptome clusters toward the claudin-low spectrum of basal-like (BL) breast cancers [6], and activation of the WNT signalling pathway is a common feature [7], with 41% (11/27) tumours exhibiting mutations in β -catenin (*CTNNB1*), *APC* or *WISP3* [8]. Early genomic

analysis using array comparative genomic hybridisation demonstrated that MBCs show greater genomic instability than IC-NST [9]. The presence of mutations in *PIK3CA* in nearly 50% of cases suggests that they are not only different to IC-NST, but may also represent a unique subgroup, as the *PIK3CA* mutation frequency in other BL/TN breast cancers is only ~8% [9]. Apart from *PIK3CA*, exome and targeted sequencing studies have revealed frequent mutations in *TP53* (69%), *PIK3R1* (11%), *ARID1A* (11%), *FAT1* (11%) and *PTEN* (11%) [7, 10], and possibly also *NOTCH* and *MCL1* [11]. Interestingly, a very high frequency (97.5%; 39/40) of A14V mutations in ribosomal protein L39 has also been reported in a CAST PCR study [12], which may have implications for the therapeutic application of nitric oxide synthase inhibitors.

There are fragmented pieces of evidence that distinct morphological features are clinically relevant. For example, Piscuoglio *et al* [10] found little variation between the spindle, chondroid and squamous morphologies (n=17) in terms of copy-number alteration, though transcriptome differences were noted in the spindle types, especially of epithelial-to-mesenchymal (EMT) genes. In a series of eight MBC, exome sequencing of separate carcinomatous and metaplastic elements revealed striking similarity, implying shared genomic ancestry [13]. Two sufficiently powered studies have so far examined survival outcomes, with the findings associating mixed morphology with the shortest DFS [14], and matrix-producing and squamous phenotypes with favourable outcome [15].

To reconcile these and other emerging research findings with clinical diagnostic practice, in 2012 the World Health Organisation (WHO) breast working group [1] published an MBC sub-classification (Table 1) to supersede the previous system, which was subjective and not universally adopted. The WHO acknowledged that the new categories were purely descriptive, but at least accurately represent discrete morphological features observed in MBC, and provided a temporary framework that can be refined in the future according to demonstrated relationships with molecular drivers and relevance to clinical management. Ascribing clinical significance to this classification through statistically sound study design is challenging because MBC samples are relatively rare in individual institutions. We therefore created the Asia-Pacific Metaplastic Breast Cancer (AP-MBC) Consortium, and collated a cohort of 347 metaplastic breast cancers from 17 hospitals across Australia and South-East Asia (supplementary material, Table S1). Here, we present detailed analysis of morphology, prognosis and molecular features in one of the largest MBC cohorts to date.

Methods

Clinical cohorts

Human research ethics committees approved the use of all clinical samples (The University of Queensland (2005000785) and Royal Brisbane and Women's Hospital (2005/022)). supplementary material, Table S1 details the origins of the cohort, as provided by the AP-MBC consortium. The Queensland Follow Up (QFU) cohort [16, 17] was also employed in this study as a cross-sectional control cohort. It comprises 527 non-metaplastic breast tumours sampled in duplicate on tissue microarrays, with detailed clinico-pathological annotation and median follow-up of 21.4 years (censored cases).

Pathology review and Immunohistochemistry

The morphological categorisation is considered as per the WHO guidelines (Table 1; [1]). Whole sections of each case were reviewed centrally by the pathology committee (GH, RY, GT, DP, RP, NP, PHT, SO, SL), either on AP-MBC consortium review days using a multi-headed microscope, or through digital file sharing following Aperio scan (40x objective magnification; Leica Biosystems, Mt Waverly, Victoria, Australia) of haematoxylin and eosin (H&E)-stained slides. For immunohistochemistry (IHC) analysis, antibodies were incubated with whole tissue sections at room temperature for 1 h; the panel of antibodies was selected to encompass a range of differentiation markers (supplementary material, Table S2). The MACH 1 Universal HRP-Polymer kit (Biocare Medical, Concord, CA, USA) was used for detection. A positive stain was defined by the presence of $\geq 5\%$ of tumour cells displaying unequivocal staining. An IHC score was calculated for

NF1 staining, whereby the intensity of staining (0 = no staining, 1 = weak staining, 2 = moderate staining, 3 = strong staining) was multiplied by the percentage of tumour cells stained (giving a score between 0 (0 x 0%) and 300 (3 x 100%).

Tumour microdissection and DNA extraction

Freshly cut formalin-fixed, paraffin embedded (FFPE) tissue sections were H&E-stained to identify tumour-rich areas and 10 µm thick tissue sections were cut onto plain slides. The tissue was deparaffinised, stained with Nuclear Fast Red and dehydrated through a graded ethanol series. Microdissection of tissue samples was performed under a stereomicroscope using a fine needle, and tissue was digested in ATL buffer (Qiagen, Chadstone, Victoria, Australia) with proteinase K in a shaking incubator at 55°C for three days, with proteinase K replenished daily to ensure complete tissue digestion. The DNA was extracted using the QIAmp DNA Micro Kit (Qiagen) following the manufacturer's instructions.

Whole exome sequencing

Whole exome sequencing was performed using the Illumina Nextera Rapid Capture Exome Kit (Illumina, Scoresby, Victoria, Australia) following the manufacturer's instructions, and libraries sequenced on an Illumina NextSeq. Sequence analysis tools are available through <https://sourceforge.net> and <https://github.com>. Sequence data was adapter trimmed using cutadapt (version 1.8.1), aligned to the GRCh37 assembly using BWA-MEM (v0.7.12-r1039) and converted to coordinate sorted lane aligned read group BAM (binary alignment mapping) files

with samtools 1.1 (using htlib 1.1; <http://www.htlib.org>). Sequence reads were filtered to include only those with an alignment length >34 or second in a pair and mapped as a proper pair; less than 3 mismatches to the reference genome and not a PCR duplicate. Thirty pairs of tumour and normal met a minimal cut off of 10x coverage, out of a possible 79 pairs sequenced. Point mutations were called using the Genome Analysis Toolkit (GATK) [18] and qSNP [19]. High confidence variants were those called by both GATK and qSNP. Variant positions had to be covered by 8 or more reads in the tumour, 12 or more reads in the normal and with the variant present in 4 or more reads with at least 4 novel starts. Large numbers of overlapping reads were observed due to the short DNA fragments in FFPE source material and custom variant calling filters were created to remove duplicate evidence due to the overlapping reads from the high confidence calls. Only high confidence variants that pass this filter were used for further analysis. Variants were annotated with population frequencies from dbSNP (v141; <https://www.ncbi.nlm.nih.gov/projects/SNP/>), ExAC (<http://exac.broadinstitute.org>; accessed July 2018) and gnomAD [20] (allele frequency <1%; accessed July 2018), and gene consequence using SnpEff 4.0e (build 2014-09-13). Genes that were significantly mutated were identified using OncodriveFM through Intogen [21, 22]; breast cancer driver genes defined as per the expanded list in [23].

Statistical analyses

Statistical analyses were performed using GraphPad Prism v7.0 (GraphPad, Inc. La Jolla, CA, USA). Kaplan-Meier curve survival associations were measured with log-rank analysis. Relationships

were assessed using either Chi square or Fisher's exact tests depending on numbers of variables. Independent datasets were used for comparison [24–26].

Results

Morphology and survival

We performed detailed clinical and morphological review (Table 2) and as shown in Figure 1A, subsets of this cohort were assessed for the expression of immunohistochemical markers, survival outcomes and subjected to exome sequencing. Considering the current diagnostic guidelines [1], the most prevalent phenotypes in the cohort were Mixed Metaplastic (WHO_1; n=251, 72.4%), pure squamous (WHO_4; 15.7%) and pure spindle (WHO_5; 8.1%; Figure 1B), accounting for 96% of all cases. Within the mixed metaplastic group there were 32 combinations of morphologies, including 12 uniquely represented (supplementary material, Figure S1). Of all the cases presenting with spindle features, 15% were classified as pure spindle (WHO_5), while a greater proportion of those presenting with squamous features (26%) were pure squamous (WHO_4), as opposed to being mixed with features of that morphology type. Considering the three largest groups, Mixed metaplastics (WHO_1) together with the pure spindle phenotype (WHO_5) had a poorer breast cancer specific survival (BCSS) over 10 years than the pure squamous (WHO_4) (Figure 1C). Statistical assessment of the clinico-pathological features showed a significant relationship between WHO type and grade, with WHO_4 tumours more likely to be Grade 2 (Figure 1D; Chi square $P=0.0278$). A trend towards a survival advantage in grade 2 tumours was noted for the pure WHO_4 and WHO_5 groups (supplementary material, Figure S2). We found no significant association between WHO type and size (Figure 1E), size and grade (Figure 1F), or WHO type and age at diagnosis (Figure 1G). The features of the metaplastic cohort showed limited similarity in distribution when compared to a cohort of grade-2 and -3 IC-NST from the Queensland Follow Up

(QFU) set [16, 17] (Table 2). In fact, the only criteria without significant differences were age at diagnosis and presence of *in situ* disease and lymphovascular invasion. In the 119 AP-MBC cases with co-occurring DCIS (Table 2 and supplementary material, Table S3), the phenotype was most commonly mixed (66%) featuring solid and/or ductal growth patterns but was significantly enriched for high-grade DCIS ($P=5.67E-06$) compared to QFU set. The BCSS analyses within the AP-MBC cohort (Figure 1H–L), showed a significant relationship only between increasing tumour size and outcome (Figure 1J; $P=0.0042$). There was a notable absence of prognostic significance for lymphocytic infiltrate, considering the significant difference in infiltrate between the AP-MBC cases and the IC-NST cohort from the QFU (Table 2; $P<1E-15$).

Comparing the BCSS of patients' presenting with pure squamous tumours (WHO_4), with those WHO_1 mixed tumours with squamous features (Figure 2A), there was a trend toward a better outcome for the pure WHO_4. However, comparing pure spindles (WHO_5) with those WHO_1 demonstrating spindle features (Figure 2B), revealed no difference in outcome. Within the WHO_1 mixed category, most cases presented with 2 phenotypes (64.9%; Figure 2C); 12.5% of the cases had 4 or more phenotypes noted. Indeed, the presence of ≥ 4 phenotypes conferred a significantly lower BCSS ($P=0.0068$, Figure 2C) than those cases with 2 or 3 morphologies. The most common morphology in these mixed tumours was carcinoma (IC-NST; 74%), followed by squamous-, then spindle- features (Figure 2D). Figure 2E shows the phenotypic co-occurrence presenting in each WHO_1 case, and the likelihood of phenotypes presenting together is measured in Figure 2F. For instance, across the WHO_1 cohort, very few cases present chondroid and squamoid

differentiation concurrently, with a significant inverse correlation ($P=1E-11$). Indeed, squamoid differentiation is also unlikely to present with either rhabdoid or osseous features. Spindle features are significantly less likely to be seen with carcinoma-NST ($P<1E-10$) than squamous features. A subset of cases with frequent large, bizarre pleomorphic cells was identified (Figure 2G). These tumours have a median size of 49 mm, and an extreme level of pleomorphism, being 'under' scored with a pleomorphism score of 3. When compared to the rest of the MBC cohort, there is a strong positive association between the presence of these cells and spindle features (15/16; $P=0.022$), however only 2/15 were WHO_5 pure spindle type; and an inverse correlation with osseous differentiation (0/16).

E-cadherin in MBC

E-cadherin expression was categorised as normal, negative or aberrant (incomplete linear/punctate staining at the membrane or diffuse cytoplasmic staining, supplementary material, Figure S3). Considering all morphologies, Figure 3A shows that the squamous regions were significantly enriched for normal or aberrant expression of E-cadherin, while spindle regions were more likely to be negative ($P<9E-14$). Figure 3B shows the E-cadherin expression across the multiple morphologies within a case, with spindles more likely to be lacking E-cadherin expression, while the squamous cells express E-cadherin aberrantly. Kaplan Meier (KM) analysis demonstrated a trend towards a survival advantage in those patients with aberrant E-cadherin, rather than a complete loss of expression (Figure 3C), consistent with the overall better prognosis of squamous tumour types. Comparing E-cadherin expression in MBC to other cancers, there was a highly

significant ($P=1.3E-14$) enrichment for aberrant and negative staining in MBCs ($n=104$), as compared to the QFU cohort (excluding MBC; $n=453$), and also to the triple negative tumours of the QFU cohort (excluding MBC; $n=62$; Figure 3D). While few AP-MBC cases presented without vimentin expression ($n=6$), 5 of these showed aberrant E-cadherin expression, and one showed combined normal and aberrant staining (Figure 3E; supplementary material, Figure S3F–I shows a representative case with E-cadherin and vimentin positivity within the same squamous foci).

Immunohistochemical phenotypes

Immunohistochemical staining was performed on 165 cases of MBC (Figure 4A,B). Contingency analyses show the WHO_1 with spindle features are significantly more likely to express CK19 than WHO_5 pure spindle tumours (Figure 4C); WHO_1 with squamous features and WHO_4 pure squamous tumours showed a significant difference in expression of vimentin and CK19 (Figure 4D); WHO_1 with chondroid features in the absence of other mesenchymal features (rhabdoid and/or osseous) showed the higher, but not significant expression of CK5/6 and CK14, while those presenting with rhabdoid and/or osseous features in the absence of chondroid were more likely to express EGFR, p63 and SMA (Figure 4E). Individual WHO type data are shown in supplementary material, Figure S4. The co-expression of markers is detailed in Figure 4F which demonstrates that the WHO_4 and 5 cases are more biologically diverse than might be expected considering they are single morphological groups. Staining for additional markers was performed on a restricted subset of cases due to limited availability of tissue ($n=65$). Well over half of the cases stained showed expression of CD44, S100, GCR, CD10, but desmin, a muscle specific intermediate filament, was

expressed by only 12% of samples (supplementary material, Figure S5A,B). No significant differences were observed in expression of these markers between squamous containing WHO_1 and pure WHO_4; or between WHO_1 spindles and pure WHO_5, as determined by a Chi square test (supplementary material, Figure S5C). Surprisingly, no association was found between expression of the stem cell marker CD44 and the number of morphologies (and thus plasticity of the tumour) in WHO_1 cases (supplementary material, Table S4).

Marker expression and survival

We considered the relationship between biomarker expression and breast-cancer specific survival within the AP-MBC cohort (Figure 5). Tumours positive for AE1/3 were associated with a significantly improved prognosis compared to those that were negative ($P=0.0072$). Conversely, EGFR expression associated with poor survival outcomes ($P=0.0111$). Of the small number of cases ($n=11$) which were negative for AE1/3 and positive for EGFR, the 5 recorded deaths were within 2 years of diagnosis. Expression of both vimentin and p63 (3+; intense staining) trended towards an association with poor survival outcomes, however it is likely that the low numbers of vimentin negative cases and p63 3+ cases impeded statistical significance. All vimentin negative cases presented with either 1 or 2 morphologies. Additionally, the subset of samples stained for GCR (glucocorticoid receptor), CD44 (breast stem marker) and CD10 (a stromal marker) showed trends towards significant survival disadvantage in the case of high-level expression of GCR and CD44, and in reduced expression of CD10 (supplementary material, Figure S5D).

Molecular landscapes

Exome sequencing of 30 matched pairs of needle-dissected tumour and normal (tumour associated or negative lymph nodes) was performed. Due to the variability in tissue block age and origin, we encountered a broad range of sequencing quality. We observed an average of 29 somatic mutations per tumour (Figure 6A, supplementary material, Tables S5 and S6), with some cases harbouring high numbers of mutations in breast cancer driver genes (range 08). The most frequently mutated genes identified are also known to be driver genes in breast cancer: *TP53* (21/30 cases), *PIK3CA* (10/30), *PTEN* (7/30) and *NF1* (4/30) (Figure 6B). Analysis of significantly mutated genes (OncodriveFM/Intogen) confirmed the enrichment of tumour drivers (*TP53*, *PIK3CA*, *PTEN*, *NF1* and *RB1*; supplementary material, Table S7). We also report co-occurring mutations in *TP53* and *PTEN*, and in *PIK3CA* and *TP53* (Figure 5C), and a highly significant enrichment for trios of co-occurring mutations in *TP53*, *PIK3CA* and *PTEN* (n=3/30; P=0.00003). *NF1* is a known breast cancer driver gene, being altered in ~5% of breast cancers, with a somatic mutation frequency of 2.9% [25], however enrichment of *NF1* mutations in MBC is a novel finding. Four MBC cases were mutated, demonstrating a significant enrichment compared to TCGA (Figure 6D, P=0.0275). The location of the identified variants, as well as those noted in TCGA are depicted in Figure 6E. Two of the *NF1* variations occur at nucleotide sites which are reported in COSMIC as pathogenically mutated in melanoma (c.2859; Q950*), in squamous lung and colon cancer (c.1571; p.Glu524fs), and breast (c.1571; p.E524Q). No significant enrichment in specific morphology was noted for the 4 mutations, though 3/4 were identified in WHO_1 tumours (Figure 6F). *NF1* IHC was performed on a subset of samples as orthogonal validation (Figure 6G). Of the 41

MBC cases examined (Figure 5H), 2 known mutant cases (MBC_026 and MBC_095) showed weak and heterogeneous expression with an IHC score of 120 and 80 respectively. MBC_145 showed very weak staining (IHC score 10), and manual review of sequencing data identified low confidence evidence of a NF1 mutation; p.Trp1952Cys, a mutation reported in COSMIC for breast cancer.

Discussion

The prognostic significance of a range of clinico-pathological features has been reported for a small number of MBC cohorts, but a consensus has yet to be reached as to the most clinically useful. Others have shown that LN status and lymphovascular invasion (LVI) impact prognosis, but not tumour size or grade [15]; or simply that LN status predicts survival [27]. In a multivariate analysis of the SEER data, Nelson *et al* [3] reported a significant association between reduced survival time and LN positivity, ER and PR negativity, high grade and tumour size. Further, in an assessment of contemporarily treated MBC, HER2 status was shown to have a significant prognostic impact, implying a responsiveness to HER2-directed therapy [28]. In contrast, our cohort demonstrates the prognostic impact of tumour size, however we do not see significant impacts on survival made by age at diagnosis, lymph node status, grade or lymphocytic infiltrate.

We have placed a particular emphasis on the current WHO classification [1], in an effort to confirm or refine what is currently a pragmatic system. We have shown that the prognosis of a pure squamous carcinoma (WHO_4) is better than a mixed tumour with squamous features. Together with the finding that the WHO_4 group is made up of both grade-2 and 3 cancers, the data

indicates that the prognostic outcome of MBC is not purely a function of grade, hence, there is value in maintaining the WHO categories as currently published [1]. We present convincing evidence to confirm suggestions that mixed metaplastic carcinomas (WHO_1) have the worst prognosis [14] and we show that increasing numbers of morphologies present in a mixed case confer a survival disadvantage. As such, when classifying a case as a WHO_1 (mixed) metaplastic carcinoma, a pathologist should also record how many different morphologies are present.

We have demonstrated that there is significant prognostic value in enumerating the biological phenotypes present within a case. Cases that show plasticity, or 'stem-like' features, show expression of breast stem cell marker CD44, and had the worst prognosis; however the proportion of cases expressing CD44 did not increase with the number of different morphologies present, as might have been expected. Similarly, those cases negative for epithelial markers E-cadherin and cytokeratins (AE1/3) also trended towards a poorer outcome. Delving into the comparisons between pure phenotypes and those of mixed morphologies, we found a significant increase in expression of CK19 observed in the WHO_1 mixed carcinoma with a spindle component compared to the pure spindle cell carcinoma (WHO_5), which undoubtedly originates from the additional morphological components; however regardless of their origin, the expression of CK19 likely has an impact on BCSS. The comparison between pure squamous (WHO_4) and WHO_1 with squamous features highlights the heterogeneity in the EMT process of these stem-like tumours, as half of the WHO_4 (pure squamous) tumours express SMA, a somewhat unexpected result given the relative homogeneity of the 'pure' squamous phenotype. Taken together with our E-cadherin

data (aberrant in squamous cells, absent from spindles), it is clear that the EMT transition occurring at the protein level in MBC is not typical, even for the basal-like phenotype, with aberrant E-cadherin localisation being the most common phenotype in AP-MBC. Indeed, the aberrant E-cadherin data is supported by research indicating that the E-cadherin complex members (p120 and β -catenin) are not localised to the cell membrane in a small series of metaplastic breast cancers [29]. Furthermore, 10/11 pure spindle cases (WHO_5) assessed do not express E-cadherin, a finding consistent with the data with of [29, 30]. The traditional, categorical dogma of an EMT, wherein vimentin switches on while E-cadherin switches off [31], is not uniformly observed in these tumours, and indeed is likely to be different within the different morphological regions. It would be interesting to determine whether the transcriptional program of EMT is also atypical in these tumours, as parallel studies of endometrial carcinosarcomas have identified differing miRNA programs between the mesenchymal and epithelial components [32].

While early reports classed strong p63 staining as a sensitive and specific marker of MBC [33], our dataset does not support this, with only WHO_3 (fibromatosis-like) cases showing 100% positivity, although the number of cases was small (n=6); WHO_1 (mixed), WHO_4 (squamous) and WHO_5 (spindle) showed a median of 63% expression. Notably, the expression of GCR in our cohort is also significantly reduced compared with that reported previously [34].

Our exome sequencing data highlights the relative paucity of mutations across these MBC cases, and confirms the high frequency of *TP53* mutation [7, 10]. Broadly, our sequencing data

recapitulates that of recent studies [7, 10, 35, 36]. An inverse correlation between *PI3KCA* mutations and the chondroid phenotype has been documented [7, 36], and while 2/10 *PI3KCA* mutated cases showed chondroid differentiation, none of the cases were purely chondroid, nor were they specifically enriched. We also presented evidence for a significant enrichment of co-occurring mutations in *TP53*, *PTEN* and *PIK3CA* in MBC, a novel finding that counters the long-held concepts that *TP53* and *PTEN* [37], and *PIK3CA* and *PTEN* [38] mutations are mutually exclusive. Supporting our data is a recent ‘cell-of-origin’ transgenic mouse study. Melchor *et al* [39] showed that 4/10 mice with targeted *Pten* loss in luminal cells developed intermediate-grade mammary metaplastic adenosquamous carcinoma, but with 8/10 tumours showing foci of metaplastic squamous cells. By adding a *Trp53* deficiency to this genotype, a high-grade metaplastic spindle cell phenotype was evident in 6/20 transgenic mice, with intermittent spindle and squamous cells seen in all tumours.

While the relationship between *NF1* alterations and breast cancer is well-known [40–42], and indeed with aggressive breast cancers [43], we have presented novel data to support a link between *NF1* mutations and the metaplastic subtype and showed weak heterogeneous expression of *NF1* in mutant cases. This data is supported by a series of six recent case reports, each highlighting an ostensibly ‘unique’ case of metaplastic breast cancer in a Neurofibromatosis patient (*NF1*; inherited mutant *NF1* syndrome) [44-49]; 4/6 cases presented with a squamous phenotype; with the remaining 2 being spindle. A single sporadic *NF1* mutation in MBC was also reported in a squamous tumour, as part of a targeted sequencing assessment of 16 MBCs [50].

Similarly, 3/4 of the *NF1* mutations identified in the AP-MBC cohort present with squamous features. Recent phase I trials show the potential benefit of an oral selective inhibitor of MAPK kinase (MEK) 1 and 2, selumetinib, in the management of neurofibromas in NF patients [51] – it is a fascinating prospect to consider the potential benefits in *NF1* mutation-harboring MBC patients, especially in light of preclinical data in TNBC xenograft models [52].

The heterogeneity within MBC is so vast, that despite our extensive cohort size, thousands of samples would need to be reviewed centrally in order to tease out the subtleties required to make any bolder changes to the guidelines. Our data demonstrate that although the WHO guidelines were developed pragmatically, they are currently appropriate. We propose a practically minor but clinically significant modification to the guidelines: all WHO_1 mixed type tumours should have the number of morphologies present recorded, as a mechanism for refining prognosis. We also show that IHC for EGFR and cytokeratins (AE1/3) is prognostic within MBC cases, and may be more valuable than P63 in the diagnostic setting. Our data indicates that the EMT taking place in these stem-like tumours is not typical, with abundant aberrant E-cadherin in squamous foci. Furthermore, we present the first report potentially linking *NF1* mutations to the pathogenesis of metaplastic breast cancer.

Acknowledgements

We thank the patients and their families, and acknowledge the support of Metro North Hospital and Health Services in the collection of the Clinical Subject Data and Clinical Subject Materials. We thank Angelika Christ, Tim Bruxner and team at the Queensland Centre for Medical Genomics for their technical expertise. The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. KN is recipient of a Keith Boden Fellowship. SO'T is funded by the National Breast Cancer Foundation (PRAC-16-006) and the Sydney Breast Cancer Foundation. NW is an NHMRC Senior Research Fellow. Pathology Queensland – Study Education and Research Committee supported the study. The study was funded by a Cancer Australia/National Breast Cancer Foundation PdCCRS grant (APP017931).

Author contributions statement

AEMR, KN, PTS, NW and SRL conceived experiments, analysed data and drafted the manuscript. AEMR, EK, KN, ML, IG, LR, DB, JK, AR, carried out experiments. KF, CN, KH, CC collated samples and clinical data. SK, QX, OH, SW, CL, JP, JS analysed data. SRL, SO'T, RY, GH, SF, PHT, DP, RP, GT, NP, KJ, MB, LK, JB, CL, ASR, MD performed pathology review. All authors were involved in writing the paper and had final approval of the submitted and published versions.

References

1. Lakhani SR, Ellis IO, Schnitt SJ, *et al.* WHO Classification of Tumours of the Breast. 4th ed. IARC: Lyon 2012.
2. El Zein D, Hughes M, Kumar S, *et al.* Metaplastic carcinoma of the breast is more aggressive than triple-negative breast cancer: a study from a single institution and review of literature. *Clin Breast Cancer* 2017; **17**: 382–391.
3. Nelson RA, Guye ML, Luu T, *et al.* Survival outcomes of metaplastic breast cancer patients: results from a US population-based analysis. *Ann Surg Oncol* 2015; **22**: 24–31.
4. Tzanninis IG, Kotteas EA, Ntanasis-Stathopoulos I, *et al.* Management and outcomes in metaplastic breast cancer. *Clin Breast Cancer* 2016; **16**: 437–443.
5. Reis-Filho JS, Pinheiro C, Lambros MB, *et al.* EGFR amplification and lack of activating mutations in metaplastic breast carcinomas. *J Pathol* 2006; **209**: 445–453.
6. Weigelt B, Kreike B, Reis-Filho JS. Metaplastic breast carcinomas are basal-like breast cancers: a genomic profiling analysis. *Breast Cancer Res Treat* 2009; **117**: 273–280.
7. Ng CKY, Piscuoglio S, Geyer FC, *et al.* The landscape of somatic genetic alterations in metaplastic breast carcinomas. *Clin Cancer Res* 2017; **23**: 3859–3870.
8. Hayes MJ, Thomas D, Emmons A, *et al.* Genetic changes of Wnt pathway genes are common events in metaplastic carcinomas of the breast. *Clin Cancer Res* 2008; **14**: 4038–4044.
9. Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, *et al.* Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res* 2009; **69**: 4116–4124.
10. Piscuoglio S, Ng CKY, Geyer FC, *et al.* Genomic and transcriptomic heterogeneity in metaplastic carcinomas of the breast. *NPJ Breast Cancer* 2017; **3**: 48.
11. Groisberg R, Hong DS, Roszik J, *et al.* Clinical next-generation sequencing for precision oncology in rare cancers. *Mol Cancer Ther* 2018; **17**: 1595–1601.
12. Dave B, Gonzalez DD, Liu ZB, *et al.* Role of RPL39 in metaplastic breast cancer. *J Natl Cancer Inst* 2017; **109**: djw292.
13. Avigdor BE, Beierl K, Gocke CD, *et al.* Whole-exome sequencing of metaplastic breast carcinoma indicates monoclonality with associated ductal carcinoma component. *Clin Cancer Res* 2017; **23**: 4875–4884.
14. Cimino-Mathews A, Verma S, Figueroa-Magalhaes MC, *et al.* A clinicopathologic analysis of 45 patients with metaplastic breast carcinoma. *Am J Clin Pathol* 2016; **145**: 365–372.
15. Rakha EA, Tan PH, Varga Z, *et al.* Prognostic factors in metaplastic carcinoma of the breast: a multi-institutional study. *Br J Cancer* 2015; **112**: 283–289.
16. Raghavendra A, Kalita-de Croft P, Vargas AC, *et al.* Expression of MAGE-A and NY-ESO-1 cancer/testis antigens is enriched in triple-negative invasive breast cancers. *Histopathology* 2018; **73**: 68–80.
17. McCart Reed AE, Saunus J, Ferguson K, *et al.* The Brisbane Breast Bank. *Open J Bioresources* 2018; **5**: p.5.
18. McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
19. Kassahn KS, Holmes O, Nones K, *et al.* Somatic point mutation calling in low cellularity tumors. *PLoS One* 2013; **8**: e74380.
20. Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**: 285–291.

21. Gonzalez-Perez A, Lopez-Bigas N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res* 2012; **40**: e169.
22. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* 2013; **10**: 1081–1082.
23. Michailidou K, Lindstrom S, Dennis J, *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017; **551**: 92–94.
24. Forbes SA, Beare D, Boutselakis H, *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017; **45**: D777–D783.
25. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490**: 61–70.
26. Gao J, Aksoy BA, Dogrusoz U, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; **6**: pl1.
27. Zhang Y, Lv F, Yang Y, *et al.* Clinicopathological features and prognosis of metaplastic breast carcinoma: experience of a major Chinese cancer center. *PLoS One* 2015; **10**: e0131409.
28. Schroeder MC, Rastogi P, Geyer CE, Jr., *et al.* Early and locally advanced metaplastic breast cancer: presentation and survival by receptor status in surveillance, epidemiology, and end results (SEER) 2010–2014. *Oncologist* 2018; **23**: 481–488.
29. Sarrio D, Rodriguez-Pinilla SM, Hardisson D, *et al.* Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res* 2008; **68**: 989–997.
30. Carter MR, Hornick JL, Lester S, *et al.* Spindle cell (sarcomatoid) carcinoma of the breast: a clinicopathologic and immunohistochemical analysis of 29 cases. *Am J Surg Pathol* 2006; **30**: 300–309.
31. Liu F, Gu LN, Shan BE, *et al.* Biomarkers for EMT and MET in breast cancer: An update. *Oncol Lett* 2016; **12**: 4869–4876.
32. Castilla MA, Moreno-Bueno G, Romero-Perez L, *et al.* Micro-RNA signature of the epithelial-mesenchymal transition in endometrial carcinosarcoma. *J Pathol* 2011; **223**: 72–80.
33. Koker MM, Kleer CG. p63 expression in breast cancer: a highly sensitive and specific marker of metaplastic carcinoma. *Am J Surg Pathol* 2004; **28**: 1506–1512.
34. Lien HC, Lu YS, Cheng AL, *et al.* Differential expression of glucocorticoid receptor in human breast tissues and related neoplasms. *J Pathol* 2006; **209**: 317–327.
35. Joneja U, Vranic S, Swensen J, *et al.* Comprehensive profiling of metaplastic breast carcinomas reveals frequent overexpression of programmed death-ligand 1. *J Clin Pathol* 2017; **70**: 255–259.
36. Krings G, Chen YY. Genomic profiling of metaplastic breast carcinomas reveals genetic heterogeneity and relationship to ductal carcinoma. *Mod Pathol* 2018; 10.1038/s41379-018-0081-z.
37. Kurose K, Gilley K, Matsumoto S, *et al.* Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. *Nat Genet* 2002; **32**: 355–357.
38. Saal LH, Holm K, Maurer M, *et al.* PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Res* 2005; **65**: 2554–2559.
39. Melchor L, Molyneux G, Mackay A, *et al.* Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models. *J Pathol* 2014; **233**: 124–137.
40. Uusitalo E, Kallionpaa RA, Kurki S, *et al.* Breast cancer in neurofibromatosis type 1: overrepresentation of unfavourable prognostic factors. *Br J Cancer* 2017; **116**: 211–217.
41. Wallace MD, Pfeifferle AD, Shen L, *et al.* Comparative oncogenomics implicates the neurofibromin 1 gene (NF1) as a breast cancer driver. *Genetics* 2012; **192**: 385–396.
42. Nik-Zainal S, Davies H, Staaf J, *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016; **534**: 47–54.

43. Yap YS, Munusamy P, Lim C, *et al.* Breast cancer in women with neurofibromatosis type 1 (NF1): a comprehensive case series with molecular insights into its aggressive phenotype. *Breast Cancer Res Treat* 2018, 10.1007/s10549-018-4851-6.
44. Chaudhry US, Yang L, Askeland RW, *et al.* Metaplastic breast cancer in a patient with neurofibromatosis. *J Clin Imaging Sci* 2015; **5**: 17.
45. Lee HS, Jung EJ, Kim JY, *et al.* Metaplastic spindle cell carcinoma of the breast in a patient with neurofibromatosis type 1. *Breast J* 2018; **24**: 391–394.
46. Natsiopoulou I, Chatzichristou A, Stratis I, *et al.* Metaplastic breast carcinoma in a patient with Von Recklinghausen's disease. *Clin Breast Cancer* 2007; **7**: 573–575.
47. Suarez-Kelly LP, Akagi K, Reeser JW, *et al.* Metaplastic breast cancer in a patient with neurofibromatosis type 1 and somatic loss of heterozygosity. *Cold Spring Harb Mol Case Stud* 2018; **4**.
48. Vellaisamy G, Mohanty S, Rout P, *et al.* Metaplastic carcinoma of breast and Neurofibromatosis 1: a rare association. *Indian J Med Paediatr Oncol* 2017; **38**: 374–376.
49. Vivas AP, Bomfin LE, Pinto CA, *et al.* Oral metastasis of metaplastic breast carcinoma in a patient with neurofibromatosis 1. *Case Rep Oncol Med* 2014; **2014**: 719061.
50. Do K, O'Sullivan Coyne G, Chen AP. An overview of the NCI precision medicine trials-NCI MATCH and MPACT. *Chin Clin Oncol* 2015; **4**: 31.
51. Dombi E, Baldwin A, Marcus LJ, *et al.* Activity of selumetinib in Neurofibromatosis Type 1-related plexiform neurofibromas. *N Engl J Med* 2016; **375**: 2550–2560.
52. Bartholomeusz C, Xie X, Pitner MK, *et al.* MEK inhibitor selumetinib (AZD6244; ARRY-142886) prevents lung metastasis in a triple-negative breast cancer xenograft model. *Mol Cancer Ther* 2015; **14**: 2773–2781.
53. 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.

Table 1. Existing MBC sub-classification*

WHO_1	mixed metaplastic carcinoma
WHO_2	low-grade adenosquamous carcinoma
WHO_3	fibromatosis-like MBC
WHO_4	squamous cell carcinoma
WHO_5	spindle cell carcinoma
WHO_6	MBC with mesenchymal differentiation:
a	Chondroid
b	Osseous
c	Other (e.g. rhabdoid)
WHO_7	myoepithelial carcinoma

*Sub-classification following [1]

Table 2. Clinical and pathological features of the AP-MBC cohort

Feature		AP-MBC		QFU (IC-NST G2/3)		P Value*
		Number of cases	% of cases	Number of cases	% of cases	
Age	<50	57	30.3	91	32.0	0.7612
	>51	131	69.7	193	68.0	
Size	T1 (<2 cm)	51	20.4	96	44.9	2.63E-08
	T2 (2-5 cm)	146	58.4	97	45.3	
	T3 (>5 cm)	53	21.2	21	9.8	
Grade	1	2	0.6	-	-	1E-15
	2	57	16.8	161	53.8	
	3	281	82.6	138	46.2	
Tubules	2	23	6.7	44	14.8	0.0011
	3	318	93.3	254	85.2	
Pleomorphism	1	1	0.3			1E-15
	2	17	5.0	120	60.3	
	3	323	94.7	179	89.9	
Mitoses	1	55	16.1	127	43.8	1.3E-14
	2	64	18.7	53	18.3	
	3	223	65.2	110	37.9	
Lymphocytic Infiltrate	none	6	1.8	90	30.3	1E-15
	mild	214	65.2	127	42.8	
	moderate	84	25.6	66	22.2	
	brisk	24	7.3	14	4.7	
DCIS	absent	131	52.8	163	54.5	0.7306
	present	119	47.6	136	45.5	
Grade	low	2	2.3	20	9.8	5.67E-06
	Int.	9	10.3	94	45.9	
	high	76	87.4	91	44.3	
LCIS	absent	214	97.3	292	97.7	0.7838
	present	6	2.7	7	2.3	
Benign disease	absent	121	70.8			
	present	50	29.2			
Lymph nodes	Negative	101	66.4	70	47.6	0.0011
	Positive	51	33.6	77	52.4	
Lymphovascular Invasion	absent	175	78.8	217	72.8	0.1235
	present	47	21.2	81	27.2	

Circumscription	<10%	110	35.9			
	10-50%	131	42.8			
	>50%	65	21.2			
Necrosis	none	109	33.4			
	focal	123	37.7			
	extensive	94	28.8			
Biomarkers						
ER	Negative	149	83.7	100	24.3	1E-15
	Positive	29	16.3	312	75.7	
PR	Negative	147	83.5	116	39.7	1E-15
	Positive	29	16.5	176	60.3	
HER2 IHC	Negative	142	87.7	228	80.3	1E-15
	Positive	20	12.3	56	19.7	
HER2 ISH	Negative	106	90.6	232	82.6	0.0456
	Positive	11	9.4	49	17.4	

DCIS, Ductal carcinoma *in situ*; ER, oestrogen receptor, HER2, human epidermal growth factor 2; IHC, immunohistochemistry; int., intermediate; ISH, *in situ* hybridisation LCIS, Lobular carcinoma *in situ*; PR, progesterone receptor. *Italic font = Statistically significant differences

Figure Legends

Figure 1. The features of the Asia-Pacific Metaplastic Breast Cancer consortium cohort.

(A) Morphological review was performed for 347 cases; of which 165 cases had blocks of sufficient quality available for IHC; clinical follow up data was available for 160 cases and exome sequencing successfully performed on 30 tumour and normal pairs. (B) WHO_1, or mixed type MBCs were most common in the AP-MBC cohort, followed by the pure squamous (WHO_4) and pure spindle (WHO_5). (C) Kaplan-Meier (KM) breast cancer specific survival (BCSS) analysis demonstrating that WHO_1 and WHO_5 have the worst prognosis of the three prominent MBC types. The WHO_2, WHO_3 and WHO_6 subtypes occurred at a very low frequency (n=1, 7 and 5, respectively), as such were not considered. (D-G) Comparative analyses assessing associations between pathology features, as measured by Chi square or Fisher's exact test (dependent on variable count): (D) n=57 grade 2; n= 282 grade 3; (E) T1 n=51, T2 n=147, T3 n=149; (F) Grade 1/2 n=59, Grade 3 n=282; (G) ≤ 50 n=57, ≥ 51 n= 131. (H-L) Kaplan-Meier survival analysis measuring clinical or pathology feature (age, lymph node status, tumour size, tumour grade, lymphocytic infiltrate, respectively) and outcome; notably only (J) tumour size showed a relationship with BCSS, with significance assessed using a log rank test.

Figure 2. Morphology and MBC

(A) Patients with pure squamous disease (WHO_4) trended towards a better overall BCSS than mixed WHO_1 tumours with squamous elements; while patients with pure spindle WHO_5 tumours did not (B). (C) Most WHO_1 mixed MBC presented with 2 or 3 reported morphologies and the total number of morphologies present significantly impacted BCSS over 10 years ($P=0.0068$); WHO_4, 5 and 6 are pure, homogenous morphologies. (D) Carcinomatous, spindle and squamous are the morphological elements most frequently reported in WHO_1 MBCs; and the distribution of morphology types per case is depicted as a heat map in (E), with the cases along the X axis, and the number of morphologies indicated on the Y axis. (F) Contingency heat map detailing the statistically determined relationships between the different morphologies. (G) 16 cases presented with bizarre pleomorphic features as shown, which were significantly enriched in

tumours containing spindle cell elements; the bar graph depicts the morphological features of the 'bizarre' cases, as compared to those of the broader AP-MBC cohort.

Figure 3. E-cadherin expression in MBC

E-cadherin expression was categorised as normal, negative or aberrant (Ab; incomplete, punctate staining at the membrane or diffuse cytoplasmic staining). (A) Contingency analysis of E-cadherin expression showing significant enrichment for normal staining in squamous regions, and for negative staining in spindle regions; chi square $P < 1E-9$ (CA, carcinoma $n=28$; SQ, squamous $n=71$; SP, spindle $n=45$; CH, chondroid, $n=11$). (B) Heatmap depicting the relationship of E-cadherin expression between morphological regions within the same case; cases are indicated along the X axis, the phenotype indicated on the Y axis, with the expression profile noted in colour (blue, normal; grey, negative; green, aberrant; purple, normal/aberrant). Note that the section scored for this marker may not have contained all phenotypes as noted in the original assessment. (C) Kaplan-Meier survival curve analysis demonstrates that aberrant E-cadherin expression trended towards a survival advantage over those cases without detectable expression. (D) Metaplastic breast cancers were more likely than other breast cancers to have aberrant E-cadherin expression. Contingency analysis comparing the QFU cohort (excluding MBC; $n=453$), exclusively triple negative cancers from QFU (excluding MBC; $n=62$) and the AP-MBC cohort ($n=104$); Chi square $P < 1E-11$. (E) Contingency analysis of co-expression of vimentin and E-cadherin, with the proportion of cases on the y axis, and the cohort split into vimentin negative ($n=6$) and vimentin positive ($n=97$), is hampered by the low number of vimentin negative cases but shows all E-cadherin negative cases were also vimentin positive.

Figure 4. Immunohistochemical features of MBC

(A) Representative images of positive staining; scoring parameters as described in methods. (B) AP-MBC cohort summary, with the proportion of cases expressing the noted biomarkers. Contingency analyses comparing, (C) WHO_1 cases with spindle features and pure WHO_5 cases,

(D) WHO_1 cases with squamous features and pure WHO_4 cases and (E) WHO_1 tumours with chondroid features alone or in the presence of rhabdoid and/or osseous differentiation, and those without chondroid but with rhabdoid and/or osseous, n=25, 10 and 13, respectively. P values (Chi square) as indicated; *, $P<0.04$; **, $P<0.01$; ***, $P=0.001$. Note that for P63 staining, only 3+ is considered positive in the detailed WHO_5 and WHO_4 analysis. (F) Map of marker co-expression across the cohort.

Figure 5. Biomarker expression and survival in MBC.

BCSS for the expression state of AE1/3, EGFR, CK5/6, CK14, CK19, p63, SMA and vimentin was assessed using Kaplan-Meier curves. Positive expression of AE1/3 and negative expression of EGFR conferred a significant survival advantage ($P=0.0072$ and 0.011 , respectively), with 3+ expression of P63 exhibiting a trend towards poorer outcome. The lack of significant P value for vimentin is because of the low numbers of cases without detectable expression.

Figure 6. Exome sequencing identifies novel molecular changes in MBC

(A) Bar graph depicting the number of mutations identified per case (left Y axis), including the number of breast cancer driver gene mutations overlayed in red dots (right Y axis), with cases on the X axis. (B) Mutation frequency in breast cancer driver genes is depicted with number of mutations of the Y axis, and the individual driver gene as noted on the X axis. (C) Contingency analyses demonstrate the enrichment for *TP53* (21/30) and *PTEN* (7/30) mutations, but not *PIK3CA* (10/30), in the AP-MBC cohort compared to a compilation breast cancer dataset (n=3357) from METABRIC [53] and the provisional TCGA exome data. Co-mutations: *TP53/PTEN* n=2/30 versus 43/3357 TCGA; *TP53/PIK3CA* n=4/30 versus 325/3357 TCGA. Note the highly significant enrichment for trios of mutations in the AP-MBC cohort (3/30) compared to TCGA (11/4357), $P=0.00021$. (D) The *NF1* mutations identified are predominantly but not significantly enriched in cases with CA/SQ features, $P=0.0715$, as shown with a contingency heatmap with mutations

noted. (E) The location of the mutations detected in AP-MBC cases (triangles) are shown on the protein schematic with amino acids and protein domains (Ras binding and the CRAL_Trio lipophilic domain) as noted on the X axis, and number of mutations on the Y axis, also including pathogenic *NF1* mutations as reported by TCGA cBioPortal (lollipops). (F) *NF1* mutations are also significantly enriched in the AP-MBC (4/30) cohort compared to TCGA (128/3357), $P=0.0275$. (G) IHC for *NF1* orthogonally validates the sequencing data, showing a reduction of protein expression (MBC_026 (bar, 140 μm), MBC_095 and the low confidence mutation MBC_145) compared to a kidney sample (positive control). Note that MBC_026 shows the heterogeneity of staining with regions of high intensity staining, and some weak staining; the blue and red box demarcate which regions the higher power images (10X) were focussed on. (H) An IHC score was generated for each case (intensity multiplied by %) and plotted. The positive control is the green square, and known mutants are in red, and marked.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods NO

Supplementary figure legends YES

Figure S1. Frequency of morphology distribution in WHO_1 cases.

Figure S2. Survival as a function of grade in WHO_4 and WHO_5 cases.

Figure S3. Examples of E-cadherin staining

Figure S4. Overview of marker expression across the WHO categories of MBC

Figure S5. Examples of a series of 5 additional markers assessed on a subset of the AP-MBC cohort using IHC

Table S1. AP-MBC cohort origins

Table S2. Details of antibodies used for IHC

Table S3. Features of DCIS

Table S4. CD44 expression in MBC

Table S5. Exome sequencing data MAF (mutation annotation format).

Table S6. Mutation landscape and morphology features of sequenced MBC

Table S7. Intogen cancer driver gene analysis

Figure_1 rp.tif

Figure_2 rp.tif

Figure_3 rp.tif

Figure_4 rp.tif

Figure_5 rp.tif

Figure_6 rp.tif

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

McCart Reed, AE;Kalaw, E;Nones, K;Bettington, M;Lim, M;Bennett, J;Johnstone, K;Kutasovic, JR;Saunus, JM;Kazakoff, S;Xu, Q;Wood, S;Holmes, O;Leonard, C;Reid, LE;Black, D;Niland, C;Ferguson, K;Gresshoff, I;Raghavendra, A;Harvey, K;Cooper, C;Liu, C;Kalinowski, L;Reid, AS;Davidson, M;Pearson, JV;Pathmanathan, N;Tse, G;Papadimos, D;Pathmanathan, R;Harris, G;Yamaguchi, R;Tan, PH;Fox, SB;O'Toole, SA;Simpson, PT;Waddell, N;Lakhani, SR

Title:

Phenotypic and molecular dissection of metaplastic breast cancer and the prognostic implications

Date:

2019-02-01

Citation:

McCart Reed, A. E., Kalaw, E., Nones, K., Bettington, M., Lim, M., Bennett, J., Johnstone, K., Kutasovic, J. R., Saunus, J. M., Kazakoff, S., Xu, Q., Wood, S., Holmes, O., Leonard, C., Reid, L. E., Black, D., Niland, C., Ferguson, K., Gresshoff, I. ,... Lakhani, S. R. (2019). Phenotypic and molecular dissection of metaplastic breast cancer and the prognostic implications. JOURNAL OF PATHOLOGY, 247 (2), pp.214-227. <https://doi.org/10.1002/path.5184>.

Persistent Link:

<http://hdl.handle.net/11343/284674>