The clinicomolecular landscape of de novo versus relapsed stage IV metastatic breast cancer

Sean Seltzer, Mark Corrigan, Seamus O'Reilly

PII: S0014-4800(19)30983-9

DOI: https://doi.org/10.1016/j.yexmp.2020.104404

Reference: YEXMP 104404

To appear in: *Experimental and Molecular Pathology*

Received date: 16 December 2019

Revised date: 15 January 2020

Accepted date: 13 February 2020

Please cite this article as: S. Seltzer, M. Corrigan and S. O'Reilly, The clinicomolecular landscape of de novo versus relapsed stage IV metastatic breast cancer, *Experimental and Molecular Pathology*(2019), https://doi.org/10.1016/j.yexmp.2020.104404

The clinicomolecular landscape of de novo versus relapsed stage IV metastatic breast cancer

Seltzer, Sean[1], Corrigan, Mark[2] O'Reilly, Seamus[2]

[1]School of Medicine, University College Cork, Cork, Ireland.
[2]Cork Breast Research Centre, Cork University Hospital, Cork, Ireland

Corresponding author email: 116100016@umail.ucc.ie (S. Seltzer)

1. Abstract

Background. *de novo* metastatic breast cancer (dnMBC) is responsible for 6-10% of breast cancer presentations with increasing incidence and has remained resistant to detection by mammography screening. Recent publications hypothesized that in addition to poor screening uptake, the presentation of dnMBC may be due to its unfavourable biology which remains unknown at the molecular level. Here we investigated the tumour biology of dnMBC in the form of clinicopathology, genomic alterations and differential gene expression to create a comparative landscape of *de novo* versus relapsed metastatic breast cancer (rMBC). Additionally, to address the current screening limitations, we conducted a preliminary biomarker investigation for early dnMBC detection.

Methods. In this retrospective case-control study, gene expression and clinical data were accessed from the Cancer Genome Atlas (TCGA) for primary tumours of treatment-naive patients with dnMBC (n=17), rMBC (n=49), and normal tissue (n=113). The clinical and histological data were assessed categorically using Fisher's Exact Test for significance ($p<0.05$), or continuously using the Mann-Whitney Test ($p<0.05$) where appropriate. The differential gene expression analysis was performed using EdgeR's negative binomial distribution model with a false discovery rate (FDR) <0.05. The resulting gene list was analysed manually for roles in metastasis as well as ontologically using STRING-DB with FDR <0.05.

Results. dnMBCs showed improved median survival vs rMBC (36 vs. 12 months). dnMBCs were more likely to be hormone receptor positive, less likely to be triple negative with lower histological lymphocytic infiltrate. In terms of genome alterations, dnMBCs had 4-fold increased PTEN mutations and poor survival with ABL2 and GATA3 alterations. Expression-wise, dnMBCs down-regulated TNFa, IL-17 signalling, and chemotaxis, while up-regulating steroid biosynthesis, cell migration, and cell adhesion. Biomarker analysis detected pre-existing and novel breast cancer biomarkers.

Conclusion. The comparative tumour landscape revealed significant clinical, pathological and molecular differences between dnMBC and rMBC, indicating that dnMBC may be a separate biological entity to rMBC at the primary level with differing paths to metastasis. Additionally, we provided a list of potential serum biomarkers that may be useful in detecting dnMBC in its pre-metastatic window if such a window exists.

Keywords: Breast cancer; Metastasis; Gene expression; de novo; Biomarkers

2. Introduction

Breast cancer (BC) is the most common cancer in women with 1.7 million new cases per year, causing 520 000 cancer-related deaths annually [1]. The incidence of breast cancer is increasing and is estimated to reach 3.2 million new cases per year by 2050 [1] With advances in treatment and the introduction of screening programs, BC mortality has decreased by 25 – 38% [2]. However there exists a subpopulation of breast cancer patients who present with stage IV or metastatic disease at the time of diagnosis, a phenomenon referred to as *de novo* metastatic or *de novo* stage IV breast cancer (dnMBC) [3]. Cancer screening programs have reduced the incidence of metastasis at diagnosis for other malignancies such as prostate cancer by 50% from 1990 to 2010, but similar initiatives for BC have had no effect on dnMBC incidence, which is currently at 6%-10% of BC presentations, accounting for 28% of metastatic breast disease and increasing [3-6]. This discrepancy in screening outcomes has given rise to avenues of investigation into the clinical features of these patients and basic histopathological classification of their tumours [3]. Here we present the first clinicomolecular landscape of these tumours and their relapsed counterparts (rMBC).

3. Materials and Methods

3.1 Clinicopathological and gene expression data. The gene expression, genomic alteration, and clinical data from treatment-naïve, primary tumours were obtained from the Cancer Genome Atlas (TCGA) and cBioPortal [7, 8]. The expression data was processed in the form of high throughput sequencing (HTSeq) counts. Patients with dnMBC (n=17) were defined as being diagnosed with Stage IV disease. rMBC patients (n=49) were defined as patients diagnosed with Stages I to III disease whose "new tumour event" was listed as "Distant Metastasis" greater than 3 months after initial diagnosis to differentiate a true relapse from undetected *de novo* metastatic disease. Normal tissue samples were also accessed for

biomarker discovery (n=113). Tumour leukocyte infiltrate quantitation was obtained from Satlz et al., 2018 [9]. Histological data was obtained from Ping et al., 2016 [10]. Mutation and copy number data were obtained from TCGA's PanCancer Atlas. Clinicopathological data was analysed for statistical significance by Fisher's exact test ($p<0.05$) or Mann-Whitney U-test ($p<0.05$) where appropriate. Survival analysis was performed using the log-rank test ($p<0.05$).

3.2 Differential gene expression analysis. Using the EdgeR package in Rstudio, implementing the negative binomial distribution model, primary tumour HTSeq counts were input, and subsequently filtered for protein-coding genes, using the Biomart package as previously described [11, 12]. Expressed genes were defined as having at least 10 counts in at least 17 samples. The counts were normalized, dispersion estimated and differentially expressed genes (DEGs) were identified by Exact Test. The results were filtered using a false discovery rate (FDR) <0.05 and further analysed by receiver operating characteristics (ROC) and area under the curve (AUC) with $p<0.05$ as previously described using GraphPad Prism version 5 [13].

3.3 Biomarker discovery. Using the EdgeR package, a similar analysis was performed using the dnMBC sample and 113 normal tissue controls from the TCGA. Significant genes (as described above) were filtered using the secretome and serum proteome accessed from the Human Protein Atlas [14]. Biomarker sensitivity and specificity was summarized in receiver operating characteristics (ROC) curves and area under the curve (AUC) with $p<0.05$.

3.4 Molecular subtype classification. Primary tumours were classified into molecular subtypes by the PAM50 signature as previously described using the genefu package in R studio and assessed for statistical significance using Fisher's exact test [15, 16].

3.5 Protein-protein interaction and Gene ontology analysis. Using STRING-DB, the significantly regulated genes were analysed for protein-protein associations using default settings; the main network of interactors was clustered in an unsupervised manner using the Markov Clustering algorithm (MCL). Each

cluster was analysed for functional enrichment in Cytoscape v3.6.1 using STRING's enrichment plugin [17]. Significant terms were defined as FDR<0.05. Enriched terms were assessed for net log2 fold change by summation of gene expression. Concomitantly, the gene list was queried through literature search for metastatic processes and their molecular mechanisms.

3.6 Genomic alterations. Mutation data, copy number variation, and alteration-based survival data were accessed from the cBioPortal via the TCGA PanCancer Atlas. TCGA sample IDs from the downloaded tumour HTSeq counts data were matched to those in cBioPortal to obtain their corresponding genomic data which were then visualized by oncogrid [8, 18].

4. Results

4.1 Clinical data. The mean ages of dnMBCs and rMBC were 61.2 $\pm$ 4.69 (95%CI) and 56.6 $\pm$3.56 (95%CI) respectively (Table 1). African American patients were more represented in rMBC (30.61% vs. 17.65%). Caucasian patients were similarly distributed (67.35% vs. 58.85%). dnMBCs were more likely to be deceased (76.47% vs. 71.43%). In terms of diagnosis, 82.35% of dnMBCs vs 61.22% were diagnosed with invasive ductal carcinoma and 5.88% vs. 22.45% were diagnosed with invasive lobular carcinoma. None of the findings were statistically significant (Table 1). An additional table has been provided with details regarding clinical variables [see Additional file 1].

4.2 Survival analysis. From the onset of metastasis the overall survivals of patients with dnMBC vs. rMBC were 36.35 months and 12.10 months respectively (p=0.0241) with a hazard ratio (HR) of 0.503 (95%CI = 0.277 to 0.914) (Fig. 1D). rMBC patient survival by metastasis free interval (MFI) >2 years was 24.32 months and showed no statistical difference to dnMBC survival (p=0.364). Relapse patients MFI< 2 years had a median survival of 10.9 months and comparing dnMBC to this group showed a statistical significance (p=0.005) with HR = 0.290 (95%CI = 0.132 to 0.638) (Figure 1E). Lastly,

comparing the MFIs of rMBC (<2 years vs. > 2 years) showed a statistical significance (p=0.0178) with HR = 2.427 (95%CI = 1.166 to 5.054).

4.3 Pathological variables. dnMBC tumours (70.59% vs. 53.06%) were ER/PR+ (p=0.0452). They were more advanced in T staging; 29.41% vs. 4.08% are T4 (p=0.0313), more lymph node involvement; 35.29% vs 16.33% are N3 (p=0.0465), more positive margins; 47.06% vs. 6.12% (p<0.001). decreased lymphocytic infiltrate; 6.17% vs 12.32% (p=0.0361), marginally increased tumour necrosis; 2.44% vs. 1.99% (p=0.0297). Histologically, dnMBCs were less aggressive with fewer mitotic cells, more tubular structures, and lower nuclear grade. In terms of molecular subtype (Figure 1F,G), dnMBCs were more likely to be Her2+ (17.65% vs. 4.08%) and less likely to be Basal (17.65% vs. 28.57%). Other non-significant findings included tumour mass and stromal content (Table 1). An additional table is available with details regarding pathological variables [see Additional file 1].

4.4 Gene expression analysis. 74 genes were up-regulated, and 57 down-regulated. Top 10 up-regulated genes were BCHE, UGT2B4, ZFP57, CALCR, BCL2L14, ARHGAP36, GPM6A, KRT4, CYP4F8, and CDC20B. Top 10 down-regulated genes were CHGA, PCSK1, GRIA1, TRH, KCNJ16, OLFM4, HDC, PI3, SIAH3, and BMP5. A complete list of DEGs has been provided [see Additional file 2]. Top performing genes by ROC and AUC analysis included PPFIBP2, GATD3A, ARC, and PWP2 (Figure 2).

4.5 Protein-protein interaction and functional analysis. The DEG list was significantly enriched in protein-protein associations (PPI enrichment p-value =1.5e-14). In total 20 clusters were formed (Figure 3) and tested for functional enrichment summarized in Additional file 3. Cluster 1 was enriched in genes involved in cell proliferation, inflammation via IL-17 and TNF, cancer pathways, cell adhesion and apoptosis signalling. Cluster 2 and 3 were enriched in nervous system processes including neural projections, cAMP signalling and calcium signalling, synaptic vesicles transport and clatherin mediated endocytosis. Cluster 6 was enriched in steroid biosynthetic processes. Cluster 8 was enriched in cell differentiation, cell adhesion, cell migration, blood vessel morphogenesis and Wnt signalling. The

complete list of functional enrichments by cluster is available [see Additional file 3]. Mechanisms of metastasis by gene expression in dnMBC versus rMBC included up-regulation of filopodia formation, Rac1/cdc42 signalling, beta-catenin signalling and adherens junction dysregulation. In rMBC, up-regulated genes were involved in MMP and urokinase plasminogen activator expression. Similarly-regulated pathways included ERK1/2, PI3K/Akt, and FAK signalling. The complete list of genes involved in metastasis is available [see Additional file 4].

4.6 Copy number alterations and mutations. Data from cBioPortal (Figure 4) showed that dnMBCs and rMBC shared the top 2 most frequently mutated genes: TP53 (37.50% vs. 34.69%); PIK3CA (31.25% vs. 28.57%). dnMBCs were more likely to have a PTEN mutation (25.00% vs. 6.12%) as well as mutations resulting in USP32 fusion proteins (18.75% vs. 0.00%) as well as KMT2C (18.75% vs. 8.16%) and GATA3 (18.75% vs. 10.20%). dnMBCs also had increased copy number alterations, and mutations but not significantly more than rMBC. In terms of alteration-based outcomes, patients with rMBC harbouring PTEN or ARID4B alterations resulted in poor survival outcomes, with no statistical significance for dnMBC. In patients with dMBC, ABL2 and GATA3 alterations resulted in poor survival outcomes, with no significance in rMBC. Patients in both groups with TP53 or PI3KCA alterations revealed no significant survival differences. Patients with TP53 mutations were associated with increased tumour hypoxia scores across both groups. No alterations were detected in either groups for the following genes in known breast cancer syndromes, namely CHEK2, MLH1, MSH2, MSH6, PMS2, EPCAM, STK11. Three rMBC patients had mutations in NF1, three had BRCA1 mutations, one with BRCA2 mutation and one with an ATM mutation. One patient in the dnMBC group harboured a NF1 mutation. A complete list of copy number alterations and mutations used in this study is available [see Additional file 5].

4.7 Biomarker analysis. 712 genes coding for experimentally confirmed secreted proteins were significantly up-regulated in dnMBCs compared to normal tissue controls. The top 5 genes were CBLN2, MMP11, COL10A2, ISBP and CARTPT. A complete list of secreted DEGs is available [see Additional

file 6]. Top performing genes include MMP11 (AUC of 1.00), followed by COL10A1 (AUC =0.9989), SCT (AUC=0.9908), and WISP1 (AUC=0.9900) (Figure 5).

5. Discussion

5.1 Study context. dnMBC is an interesting phenomenon given its steadily increasing incidence despite mammography screening and how it seems to challenge the Halstedian paradigm of BC tumour progression [3, 5]. In this work we have conducted the first gene expression study of these tumours, establishing a preliminary molecular portrait of this disease and have shown that there are indeed significant clinical, genomic, molecular and pathological differences between dnMBC and relapsing primary tumours, indicating that dnMBC may have distinct biology. Due to this study's small sample size, largely resulting from limited primary tumour data, clinical consistency with the literature was important to improve its external validity. Significant findings that were recapitulated in this study included the increased hormone receptor positive status of dnMBC, higher nodal involvement, improved survival outcomes relative to rMBC, and the importance of the MFI for rMBC prognosis [19-21]. Non-significant congruencies were increased age of metastasis for patients with dnMBC. The main non-significant incongruencies included increased diagnosis of IDC in dnMBCs and increased frequency of African American patients in rMBCs. It should also be noted that while on average dnMBC survival outcomes are better than those in relapsed patients, recent publications have shown conflicting evidence for the role of systemic therapy in terms of patient survival between these tumour groups. One study found that in patients having undergone systemic therapy, dnMBCs have a better prognosis than rMBC regardless of MFI (greater or less than 24 months) [22]. However, another study showed similar outcomes between dnMBC and rMBC with MFI >24 months regardless of use of systemic (neo)adjuvant therapy [21]. In using samples from the TCGA, none of this study's patients received neo-adjuvant therapy and their clinical outcomes are more comparable with the findings in the latter of the aforementioned studies. Despite this study's sample size, the clinical data concur with both studies regarding the importance of the

MFI for survival in rMBC, however this study's data regarding systemic therapies after tumour resection is too limited to comment on (Table 1).

5.2 dnMBC expresses more therapeutic targets than rMBC. According to the National Institute for Health and Care Research (NICE) guidelines, in cases of advanced breast cancer, treatment is guided by tumour histology. Histology focuses on important therapeutic targets, namely the oestrogen receptor, progesterone receptor and human epidermal growth factor receptor (Her2/ Erbb2). In this study dnMBC were more likely to be HR+ and Her2+ than rMBC, thus increasing the use of endocrine therapy (Tamoxifen/ Anastrazole) and epidermal growth factor targeting therapy (Trastuzumab and Lapatinib) in this treatment naïve subgroup. Furthermore, in advanced breast cancer, both HR and Her2 are associated with improved survival, though the benefits of Her2 expression appear to be unrelated to its use as a therapeutic target [21, 23, 24]. Contrarily, rMBC was more likely to be Basal in molecular subtype and triple negative histologically which are known to be more aggressive tumours with poor survival outcomes [25].

5.3 Genomic alterations have group-specific effects on patient outcomes. In assessing the genomic landscape of these tumours, we uncovered that PTEN, a tumour suppressor phosphatase and tensin homolog, was more frequently mutated in dnMBC [26]. Since PTEN loss is a known tumourigenic event in BC with prognostic implications, we examined the effects of its alteration in each group on patient survival [27]. Interestingly, despite the increased presence of PTEN mutations and similar levels of deletions, PTEN alterations had no significant prognostic effect on patients with dnMBC, however in rMBC a PTEN alteration appeared to be a devastating prognostic event (Figure 4D). This finding led us to examine more gene alterations for group-specific survival effects and discuss their implications. In terms of the main offenders in breast cancer, TP53 and PIK3CA alterations showed no survival differences in either group. ARID4B, a gene interacting with chromatin modifying complexes and associated with metastasis in BC, was shown to be more frequently amplified in rMBC with poor survival outcomes [28]. For dnMBC patients, alterations in GATA3 and ABL2 had poor survival outcomes, while rMBC patients

were not significantly affected by these gene alterations (Figure 4D). GATA3 is a transcription factor that regulates normal breast morphology and is correlated with the expression of ER [29]. While there is debate over the role of this gene in BC, it has been hypothesized that mutations in GATA3 in ER+ tumours may alter ER turnover and enhance ER and GATA3-driven tumour growth [29]. Conversely, wild-type GATA3 expression can also repress Basal tumour progression which is a possible explanation for why we observe GATA3 mutations in rMBC with similar patient outcomes [29].

5.3 dnMBCs down-regulate immune infiltration. Transcriptomic analysis and functional enrichment of protein clusters revealed that dnMBCs down-regulate chemotaxis, TNFa, interleukin-17 (IL-17) signalling and the inflammatory response which is consistent with the significant histological finding of a 2-fold decrease in tumour-infiltrating lymphocytes (TILs). Previous studies have discussed the role of the immune response to chemotherapy in the poor outcomes of rMBC [22]. In this study, even before systemic therapy, a significant increase in TILs was observed which may indicate that even in primary lesions, patients with rMBC may be primed for a tumourigenic response to such therapies through its pro-inflammatory biology. In murine breast cancer models, it has been demonstrated that TILs, particularly T-cells that secrete IL-17, promote a pro-tumorigenic and pro-inflammatory environment that results in increased tumour proliferation, angiogenesis, and increased expression of matrix metallopeptidase 9 (MMP9) which degrades the extracellular matrix (ECM), thus promoting invasion and metastasis [30]. Interestingly, all these aforementioned factors were enriched in rMBC for Cluster 1, indicating that immune-mediated tumour progression may be a significant differentiating pathway to metastasis between these two tumour groups. And indeed, when the gene list was queried for molecular mechanisms of metastasis, rMBC showed increased proclivity for ECM degradation, targeting both urokinase plasminogen activator (uPA) and MMP expression. Additionally, MMP9 expression was found to be up-regulated in rMBC and PTEN mutations were 4 times more common in dnMBC and have been previously associated with an immune evasion phenotype in dnMBC [31].

5.4 dnMBCs up-regulate steroid biosynthesis with implications for endocrine therapy. Steroid signalling is well-characterized feature that drives ER/PR+ breast cancer [32]. In keeping with their significantly increased ER/PR+ histology, dnMBCs showed net up-regulation in cholesterol/steroid synthetic processes in Cluster 6. One of the up-regulated genes, namely HSD17B7 is known to be induced by oestrogen receptor alpha (ERa) and drives tumourigenesis through a feedforward mechanism involving the production of intratumoural oestradiol from weaker steroids oestrogen and oestrone [33]. MSMO1, another gene involved in cholesterol biosynthesis was also shown to be up-regulated in ER+ BC cell lines and whose increased expression was associated with resistance to aromatase inhibitors [34]. While dnMBCs are more likely to be ER+ than rMBC they also express genes involved in endogenous steroid production and aromatase inhibitor resistance which has implications for endocrine therapy. One study demonstrated the importance of systemic therapy (including endocrine therapy) in dnMBC survival, without which the survival benefit is lost relative to rMBC which implicates unrestrained steroid signalling in dnMBC tumour progression [22].

5.5 dnMBCs harness the cytoskeleton and disrupt cell adhesion to promote invasion and metastasis. A prominent theme among the genes involved in dnMBC metastasis mechanisms is the interconnectivity between filopodia assembly dynamics, Rac1/cdc42, beta-catenin, and adherens junctions, namely Wnt signalling. Canonical Wnt signaling leads to nuclear localization of beta-catenin which in normal epithelium is in close approximation to the cell membrane where it stabilizes cell-cell adhesion via E-cadherin [35]. Displacement of beta-catenin from adherens junctions destabilizes them and decreases epithelial integrity—an essential step to epithelial mesenchymal transformation (EMT). Additionally, Rac1 promotes the nuclear localization of beta-catenin and also regulates filopodia formation through actin dynamics in non-canonical Wnt signalling [36]. Wnt11 (up-regulated in dnMBC) has been shown to activate both canonical and non-canonical Wnt pathways, possibly combining the EMT-promoting effect of nuclear beta-catenin through canonical Wnt signalling with metastasis-promoting actin regulation via non-canonical signalling [37]. An in vitro Wnt pathway knockdown study in breast cancer cells

reinforced this concept in demonstrating that Wnt1 depletion caused actin disorganization, decreased Rac1 expression and interfered with filopodia function and decreased the cancer stem cell's migratory potential [38, 39]. This relative dependence on cytoskeletal dynamics indicates that dnMBCs may be more sensitive to cytoskeletal-targeting therapies such as taxanes which are commonly used in the treatment of advanced breast cancer and may contribute to their improved survival outcomes as supported by a previous study highlighting the importance of systemic therapy for improved dnMBC survival [22]. Conversely, rMBC tumour cells may rely more heavily on ECM degradation to propagate rather that cytoskeletal activity. Current therapies for systemic disease do not specifically target cancer cell's enzymatic degradation of the ECM, which is associated with most aggressive form of breast cancer, namely triple negative (histological) or basal (molecular) subtypes which are overrepresented in rMBC. These findings are also consistent with patient survival outcomes. Interestingly, development of treatment resistance was previously hypothesized to be the cause of poor survival in rMBC, however, in this study we have preliminary evidence that even in the primary stage, tumours of rMBC are more likely to be basal, accompanied by a less targetable mechanism of progression and metastasis compared to dnMBC.

5.6 Mammography and dnMBC. Though the tumour sizes for this study were missing, a recent study using Surveillance, Epidemiology, and End Results (SEER) data demonstrated that dnMBC primary lesions tend to be larger than stages I-III tumours which are routinely detected mammographically, indicating that tumour size is not likely to be the limiting factor regarding mammography detection [5]. Rather, it has been postulated that dnMBC can grow rapidly and metastasize between mammograms, making the disease difficult to detect, and tending to occur in populations with limited access to screening, namely, African American women, and those of low socioeconomic and educational status [5]. However, this would appear biologically discordant. Despite their rapid development, dnMBC primary lesions appear less aggressive than their rMBC counterpart, yet have significantly increased tumour T staging, indicating that despite lower mitotic bodies, lower nuclear grade, and relatively preserved tubular architecture, these primary tumours are growing sufficiently larger, and/or reaching the

skin or chest wall, relative to primary tumours of rMBC patients. An explanation for this clinicopathological discrepancy is that these tumours may have experienced a period of rapid growth followed by a plateau at the time of resection according to a logistic or Gompertz model of tumour growth [40]. This finding is intriguing, especially when matched with the increased incidence of PTEN mutations which are associated with increased tumour size, increased tumour stage, poor differentiation and poor clinical outcomes [27, 41]. Additionally, Harding and Welch have noted that while the incidence of small tumours has increased, the incidence of large tumours has not proportionately decreased, indicating that improved screening is not detecting most of these large tumours in earlier stages [42, 43]. Furthermore, Welch and colleagues noted that in countries with limited BC screening, the incidence of dnMBC is similar indicating that their size and presentation are likely reflective of their unfavourable biology [3, 5]. The question then becomes, is dnMBC an inevitable product of tumour biology, or does a window exist where these tumours can be detected and treated before metastasis? Is the answer more screening or is screening doing more harm if biology rather than early detection dictates patient survival?

5.7 dnMBCs express sensitive and specific secreted protein biomarkers. Serum biomarkers have long been sought for breast cancer and are currently in their early phase of development [44-46]. With the limitations of BC screening in the context of dnMBC, we searched for adjunct screening methods for this disease in the form of serum biomarkers. Here we produced a panel of previously confirmed secreted proteins that are sensitive and specifically expressed by dnMBC. Interestingly, some of the biomarkers discovered here have been detected in sera of patients for breast cancer, namely PLAC1, FN1 (FN), EDIL3 (DEL1), TFF1, TFF3, AGR2, AGR3, APOC1, and PTN, though they do not reach the high AUC values (>0.99) that our top-performing biomarkers exhibit [45, 46]. To further validate our candidates, protein expression studies in patient plasma will need to be undertaken. However, even if successful, the biomarkers will be of little use if outreach and healthcare access are not improved for the low socioeconomic demographic of women that constitutes dnMBC.

6. Conclusion

In this study, we established a preliminary molecular landscape of dnMBC versus rMBC primary tumours to further understand how they differ and revealed some significant biological insights. dnMBC appears to be less aggressive than rMBC despite its early metastatic potential; this is supported by patient survival, histological grading, molecular subtyping and by our molecular model. Briefly, dnMBC showed increased proclivity for cytoskeletal regulation, was more steroid dependent, and recruited fewer lymphocytes, while rMBC was more immunogenic, more likely to be triple negative and targeted the ECM more frequently. Ultimately the limitations of mammography with respect to dnMBC may be compensated for through the addition of sensitive and specific serum biomarker screening that will prompt further diagnostic imaging; the candidates we discovered in this study would require further validation in patient serum. Lastly, due to this study's small sample size, we encourage further research into the molecular properties of these primary tumours, as well as their metastatic counterparts, to further characterize the molecular differences between them which may have important implications for therapy and tumour detection.

List of Abbreviations:

AUC: area under the curve
BC: breast cancer
CNA: copy number alterations
dnMBC: de novo metastatic breast cancer
DEG: differentially expressed genes
ECM: extracellular matrix
ER: estrogen receptor alpha
HER2: human epidermal growth
HR: hazard ratio
MFI: metastasis-free interval
MMP: matrix metalloproteinase
PFI: progression-free interval
PPI: protein-protein interaction
PR: progesterone receptor
rMBC: recurrent metastatic breast cancer
ROC: receiver operant characteristics
TCGA: the cancer genome atlas

TNFa: tumour necrosis factor alpha

Declarations:

Acknowledgements

Not applicable

Ethical approval and consent to participate

Ethical approval for this study was obtained from University College Cork's Clinical Research Ethics Committee (CREC).

Availability of data and materials

All tumour data is publicly available through the TCGA portal (https://portal.gdc.cancer.gov/) and cBioPortal (https://www.cbioportal.org/study/summary?id=brca_tcga_pan_can_atlas_2018). The data produced in this study can be accessed in the "Additional files" section.

Authors' contributions

SOR and MC conceived the study and its design. SS performed the statistical analyses and gene expression studies. All authors wrote, reviewed and approved of the final manuscript.

Competing interests

The authors declare that they have no competing interests

Consent for publication

No consent was required as per CREC and TCGA's de-identification of public access data.

Funding

This study was funded by the Health Research Board of Ireland.

References:

1. Tao, Z., et al., *Breast Cancer: Epidemiology and Etiology.* Cell Biochem Biophys, 2015. **72**(2): p. 333-8.
2. Berry, D.A., et al., *Effect of screening and adjuvant therapy on mortality from breast cancer.* N Engl J Med, 2005. **353**(17): p. 1784-92.
3. Welch, H.G., D.H. Gorski, and P.C. Albertsen, *Trends in Metastatic Breast and Prostate Cancer--Lessons in Cancer Dynamics.* N Engl J Med, 2015. **373**(18): p. 1685-7.
4. Cortesi, L., et al., *Twenty-years experience with de novo metastatic breast cancer.* Int J Cancer, 2015. **137**(6): p. 1417-26.
5. Heller, D.R., et al., *Why Has Breast Cancer Screening Failed to Decrease the Incidence of de Novo Stage IV Disease?* Cancers (Basel), 2019. **11**(4).
6. Mariotto, A.B., et al., *Estimation of the Number of Women Living with Metastatic Breast Cancer in the United States.* Cancer Epidemiol Biomarkers Prev, 2017. **26**(6): p. 809-815.
7. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
8. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
9. Saltz, J., et al., *Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images.* Cell Rep, 2018. **23**(1): p. 181-193 e7.
10. Ping, Z., et al., *A microscopic landscape of the invasive breast cancer genome.* Scientific reports, 2016. **6**: p. 27545-27545.
11. Smedley, D., et al., *BioMart--biological queries made easy.* BMC Genomics, 2009. **10**: p. 22.
12. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
13. Parodi, S., et al., *ROC curves are a suitable and flexible tool for the analysis of gene expression profiles.* Cytogenet Genome Res, 2003. **101**(1): p. 90-1.
14. Thul, P.J. and C. Lindskog, *The human protein atlas: A spatial map of the human proteome.* Protein Sci, 2018. **27**(1): p. 233-244.
15. Gendoo, D.M., et al., *Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer.* Bioinformatics, 2016. **32**(7): p. 1097-9.
16. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* J Clin Oncol, 2009. **27**(8): p. 1160-7.
17. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
18. Cooper, L.A., et al., *PanCancer insights from The Cancer Genome Atlas: the pathologist's perspective.* J Pathol, 2018. **244**(5): p. 512-524.
19. Dawood, S., et al., *Survival differences among women with de novo stage IV and relapsed breast cancer.* Ann Oncol, 2010. **21**(11): p. 2169-74.

20. Yamamura, J., et al., *The Difference in Prognostic Outcomes Between De Novo Stage IV and Recurrent Metastatic Patients with Hormone Receptor-positive, HER2-negative Breast Cancer.* In Vivo, 2018. **32**(2): p. 353-358.
21. Lobbezoo, D.J., et al., *Prognosis of metastatic breast cancer: are there differences between patients with de novo and recurrent metastatic breast cancer?* Br J Cancer, 2015. **112**(9): p. 1445-51.
22. Shen, T., et al., *Prognostic outcomes in advanced breast cancer: the metastasis-free interval is important.* Hum Pathol, 2017. **70**: p. 70-76.
23. Yardley, D.A., et al., *Treatment patterns and clinical outcomes for patients with de novo versus recurrent HER2-positive metastatic breast cancer.* Breast Cancer Res Treat, 2014. **145**(3): p. 725-34.
24. Ren, Z., et al., *Prognostic factors in advanced breast cancer: Race and receptor status are significant after development of metastasis.* Pathol Res Pract, 2016. **212**(1): p. 24-30.
25. Hudis, C.A. and L. Gianni, *Triple-negative breast cancer: an unmet medical need.* Oncologist, 2011. **16 Suppl 1**: p. 1-11.
26. Steck, P.A., et al., *Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers.* Nat Genet, 1997. **15**(4): p. 356-62.
27. Li, S., et al., *Loss of PTEN expression in breast cancer: association with clinicopathological characteristics and prognosis.* Oncotarget, 2017. **8**(19): p. 32043-32054.
28. Winter, S.F., et al., *Allelic variation and differential expression of the mSIN3A histone deacetylase complex gene Arid4b promote mammary tumor growth and metastasis.* PLoS Genet, 2012. **8**(5): p. e1002735.
29. Takaku, M., S.A. Grimm, and P.A. Wade, *GATA3 in Breast Cancer: Tumor Suppressor or Oncogene?* Gene Expr, 2015. **16**(4): p. 163-8.
30. Benevides, L., et al., *IL17 Promotes Mammary Tumor Progression by Changing the Behavior of Tumor Cells and Eliciting Tumorigenic Neutrophils Recruitment.* Cancer Research, 2015. **75**(18): p. 3788.
31. Jain, E., et al., *Abstract PD9-03: The genomic landscape of de novo metastatic breast cancer (MBC).* Cancer Research, 2019. **79**(4 Supplement): p. PD9-03.
32. Foster, P.A., *Steroid metabolism in breast cancer.* Minerva Endocrinol, 2008. **33**(1): p. 27-37.
33. Shehu, A., et al., *The stimulation of HSD17B7 expression by estradiol provides a powerful feed-forward mechanism for estradiol biosynthesis in breast cancer cells.* Molecular endocrinology (Baltimore, Md.), 2011. **25**(5): p. 754-766.
34. Simigdala, N., et al., *Cholesterol biosynthesis pathway as a novel mechanism of resistance to estrogen deprivation in estrogen receptor-positive breast cancer.* Breast cancer research : BCR, 2016. **18**(1): p. 58-58.
35. Brembeck, F.H., M. Rosario, and W. Birchmeier, *Balancing cell adhesion and Wnt signaling, the key role of beta-catenin.* Curr Opin Genet Dev, 2006. **16**(1): p. 51-9.
36. Schlessinger, K., A. Hall, and N. Tolwinski, *Wnt signaling pathways meet Rho GTPases.* Genes Dev, 2009. **23**(3): p. 265-77.
37. Tao, Q., et al., *Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos.* Cell, 2005. **120**(6): p. 857-71.
38. Jang, G.-B., et al., *Blockade of Wnt/β-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype.* Scientific Reports, 2015. **5**: p. 12465.
39. Koval, A. and V.L. Katanaev, *Dramatic dysbalancing of the Wnt pathway in breast cancers.* Sci Rep, 2018. **8**(1): p. 7329.

40. Gerlee, P., *The Model Muddle: In Search of Tumor Growth Laws.* Cancer Research, 2013. **73**(8): p. 2407-2411.
41. Zhang, H.-Y., et al., *PTEN mutation, methylation and expression in breast cancer patients.* Oncology letters, 2013. **6**(1): p. 161-168.
42. Harding, C., et al., *Breast Cancer Screening, Incidence, and Mortality Across US Counties.* JAMA Intern Med, 2015. **175**(9): p. 1483-9.
43. Welch, H.G., et al., *Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness.* N Engl J Med, 2016. **375**(15): p. 1438-1447.
44. Kazarian, A., et al., *Testing breast cancer serum biomarkers for early detection and prognosis in pre-diagnosis samples.* British journal of cancer, 2017. **116**(4): p. 501-508.
45. Yuan, H., et al., *PLAC1 as a serum biomarker for breast cancer.* PLoS One, 2018. **13**(2): p. e0192106.
46. Loke, S.Y. and A.S.G. Lee, *The future of blood-based biomarkers for the early detection of breast cancer.* Eur J Cancer, 2018. **92**: p. 54-68.

Additional Files

Additional file 1: Patient and tumour characteristics. Table consisting of TCGA patient IDs clinical variables including age, gender, ethnicity and survival, as well as pathological variables including, diagnosis, histochemistry, PAM50, etc.

Additional file 2: Differentially expressed genes dnMBC vs. rMBC: Complete table of 131 differentially expressed genes, showing the gene's Ensembl ID, followed by gene name, log fold change, log(counts per million), p-value and FDR.

Additional file 3: Gene ontology analysis. Table of 20 gene clusters with additional details regarding enriched GO terms, number of genes involved, FDR, net regulation, etc.

Additional file 4: Mechanisms of metastasis. Table of DEGs with known involvement in mechanisms of metastasis, showing gene name, category, mechanism, reference and tabulation of the genes below.

Additional file 5: CNA and mutations. Table providing details of copy number alterations and mutations in dnMBC and rMBC primary tumours.

Additional file 6: Differentially expressed genes dnMBC vs. normal tissue: Complete table of 1228 differentially expressed, secreted genes, showing the gene's Ensembl ID, followed by gene name, whether that gene is present in normal human plasma, log fold change, log(counts per million), p-value and FDR.

Figure Legends

Figure 1: Clinicopathological characteristics of dnMBC vs. rMBC. (A-C) PAM50 molecular classification of tumours showing increased Her2 and less Basal tumours in dnMBC. (B) Patient age boxplot showing similar age distributions. (D,E) Patient overall survival from onset of metastasis indicating that dnMBC patients have a better prognosis. (F,G) Tumour histology from the TCGA showing low grade versus high grade tumours respectively and (H-K) histological characteristics that contribute to histological grade showing that dnMBCs have a lower grade. (L) Plot of lymphocytic infiltrate tumour fraction with decreased infiltration in dnMBC. (M) Tumour T-staging profile with significantly elevated T staging in dnMBC.

Figure 2: Differential gene expression analysis of dnMBC vs. rMBC. (A) Hierarchical clustering of top performing DEGs showing clustering of dnMBC. (B-E) LogCPM expression and AUC performance of the top 4 DEGs which have unknown roles in breast cancer. (F,G) Principal component analysis of the top performing DEGs showing segregation of tumours.

Figure 3: Gene ontology analysis and comparative metastasis mechanisms. (A) Main clusters (1-20) of core DEGs in PPI-interaction network. (B) Gene ontology terms by cluster with FDR and log fold change showing changes in inflammatory response, chemotaxis, cell adhesion and steroid synthesis. (C)

Molecular model of metastasis highlighting main differences by gene expression, including filopodia assembly, MMP activation/ expression, Rac1 signalling and cell adhesion. (D) Mechanisms of metastasis by gene expression.

Figure 4: Oncogrid and alteration-specific survival of dnMBC vs rMBC. (A) Oncogrid displaying tumour mutations, copy number alterations and hypoxia scores. (B) Comparison of most frequently mutated genes, namely TP53, PIK3CA, PTEN, USP32, and KMT2D. (C) Comparison of copy number variations between groups. (D) Differing effects of gene alterations in PTEN, ARID4B, ABL2, and GATA3 on overall patient survival in dnMBC and rMBC.

Figure 5: Biomarker discovery analysis. (A) Unsupervised hierarchical clustering of DEGs between Normal breast tissue and dnMBC showing clustering of tumour and normal tissue. (B-E) Expression and ROC analysis for top performing biomarker candidates: MMP11, COL10A1, SCT, and WISP1. (F,G) Principal component analysis of DEGs showing tumour segregation from normal tissue.

Table1 Clincopathological characteristics

Characteristics	dnMBC (n = 17) Number	dnMBC (n = 17) Percentage	rMBC (n = 49) Number	rMBC (n = 49) Percentage	Exact test p-value	T test p-value
Age					0.147	--
<50	2	11.76%	17	34.69%		
50–64	11	64.71%	20	40.82%		
65+	4	23.53%	12	24.49%		
Age at metastasis					--	0.436
	61.24	--	58.88	--		
Year of diagnosis					0.65	
1990-2000	2	12.50%	3	6.12%		
2000-2010	3	18.75%	12	24.49%		
>2010	12	75.00%	33	67.35%		
Unknown	0	0.00%	1	2.04%		
Diagnosis					0.364	--
IDC	14	82.35%	30	61.22%		
ILC	1	5.88%	11	22.45%		
Mucinous	0	0.00%	1	2.04%		
Mixed	1	5.88%	3	6.12%		
Unknown	1	5.88%	4	8.16%		
Tumor stage					n/a	--
I	--	--	3	6.12%		
II	--	--	22	44.90%		
III	--	--	24	48.98%		
Tumour Size					**3.13E-02**	--
T1	1	5.88%	7	14.29%		
T2	6	35.29%	28	57.14%		
T3	4	23.53%	11	22.45%		
T4	5	29.41%	2	4.08%		
Unknown	1	5.88%	0	0.00%		
Node status					**4.65E-02**	--
N0	0	0.00%	7	14.29%		
N1	3	17.65%	24	48.98%		
N2	4	23.53%	10	20.41%		
N3	6	35.29%	8	16.33%		
Unknown	4	23.53%	0	0.00%		
Histology					0.22	--
ER+/PR+	12	0.67	26.00	0.51		
ER-/PR-	0	0.00	9.00	0.18		
Her2+	1	0.06	2.00	0.04		
Triple neg.	1	0.06	7.00	0.14		
Unknown	4	0.22	7.00	0.14		
ER/PR status					**4.52E-02**	--

Positive	12	70.59%	26	53.06%		
Negative	1	5.88%	16	32.65%		
Unknown	4	23.53%	7	14.29%		
Her2 Status					1	--
Positive	1	5.88%	2	4.08%		
Negative	12	70.59%	26	53.06%		
Unknown	4	23.53%	21	42.86%		
Histological Grade					0.0856	
Grade I	1	5.88%	3	6.12%		
Grade II	6	35.29%	4	8.16%		
Grade III	5	29.41%	20	40.82%		
Unknown	5	29.41%	22	44.90%		
Tumor Mass					0.099	--
<300	3	17.65%	14	28.57%		
300-600	8	47.06%	24	48.98%		
600-900	5	29.41%	3	6.12%		
>900	1	5.88%	7	14.29%		
Unknown	0	0.00%	1	2.04%		
PAM50 type					0.481	--
Luminal A	4	23.53%	12	24.49%		
Luminal B	6	35.29%	17	34.69%		
Her2	3	17.65%	2	4.08%		
Basal	3	17.65%	14	28.57%		
Normal	1	5.88%	4	8.16%		
Tumour Infiltrate						
Leukocytes	--	16.05%	--	23.74%	--	0.165
Macrophages	--	8.41%	--	9.74%	--	0.351
Lymphocytes	--	6.17%	--	12.32%	--	**3.61E-02**
Neutrophils	--	0.05%	--	0.06%	--	0.588
Mast cells	--	0.93%	--	1.15%	--	0.394
Dendritic cells	--	0.50%	--	0.47%	--	0.101
Eosinophils	--	0.00%	--	0.00%	--	n/a
Margin status					**7.41E-05**	--
Positive	8	0.47	3.00	0.06		
Negative	6	0.35	42.00	0.86		
Close	2	0.12	2.00	0.04		
Unknown	1	0.06	2.00	0.04		
Tumour necrosis					--	**2.97E-02**
% necrosis		2.44%		1.99%		
Tumor stromal cells					--	0.229
% stromal		21.68%		16.19%		
Site of Metastasis					--	--
Bone	2	10.53%	33	63.46%		
Brain	1	5.26%	2	3.85%		

Lung	1	5.26%	11	21.15%		
Liver	1	5.26%	1	1.92%		
Skin	0	0.00%	2	3.85%		
Cervical node	0	0.00%	1	1.92%		
Mediastinal nodes	0	0.00%	2	3.85%		
Unknown	14	73.68%	0	0.00%		
Systemic therapy					--	--
Neo-adjuvant	0	0.00%	0	0.00%		
Adjuvant	3	17.65%	2	4.08%		
None	1	5.88%	0	0.00%		
Unknown	13	76.47%	47	95.92%		
Ethnicity					0.664	--
Black	3	17.65%	15	30.61%		
White	10	58.82%	33	67.35%		
Unknown	4	23.53%	1	2.04%		
Vital status					0.762	--
Living	4	23.53%	14	28.57%		
Deceased	13	76.47%	35	71.43%		
MFI						
<6 months	--	--	3	6.12%		
6 months - 2 years	--	--	24	48.98%		
>2 years	--	--	22	44.90%		

Author statement

S. O'Reilly: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration. **S.Seltzer**: Methodology, Software, Visualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing. **M. Corrigan**: Conceptualization, Methodology, Supervision, Writing - Original Draft, Writing - Review & Editing.

Highlights:

- From the onset of metastasis, de novo metastatic breast cancer (dnMBC) patients have increased median survival compared to their relapsed counterparts (rMBC).
- Relative to rMBC, dnMBC primary tumours display an immune evasion phenotype in their transcriptomes with significantly reduced tumour infiltrating lymphocytes histologically.
- Genomic alterations in PTEN, GATA3, ABL2 and ARID4B have differential effects on patient survival in dnMBC vs. rMBC.
- dnMBC tumours express sensitive and specific biomarkers that may be detectable in patient serum.

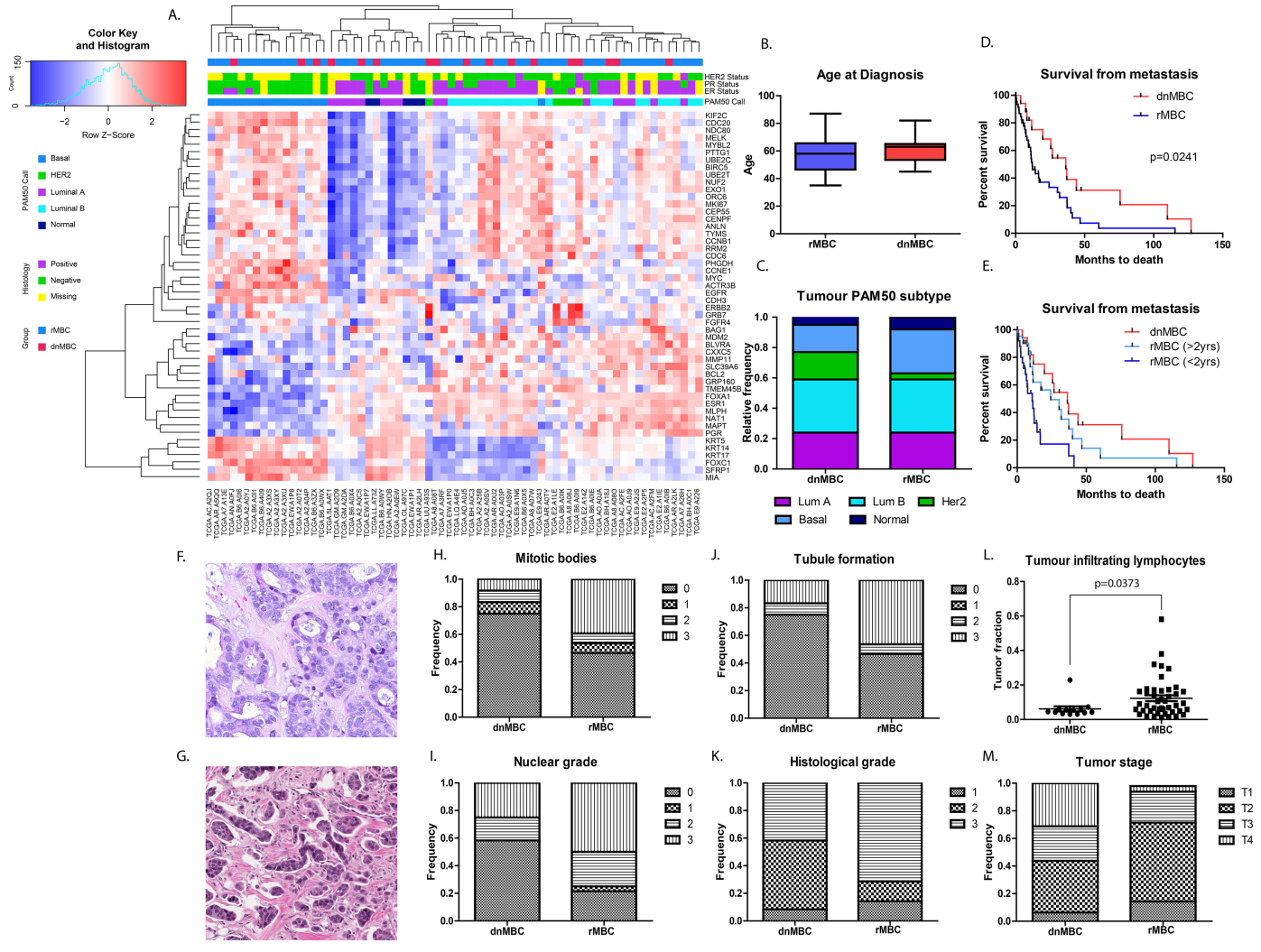

A.
Color Key and Histogram
Count
Row Z-Score
HER2 Status
PR Status
ER Status
PAM50 Call
Basal
HER2
Luminal A
Luminal B
Normal
Positive
Negative
Missing
rMBC
dnMBC
B.
Age at Diagnosis
Age
rMBC
dnMBC
C.
Tumour PAM50 subtype
Relative frequency
Lum A
Lum B
Her2
Basal
Normal
D.
Survival from metastasis
dnMBC
rMBC
p=0.0241
Percent survival
Months to death
E.
Survival from metastasis
dnMBC
rMBC (>2yrs)
rMBC (<2yrs)
Percent survival
Months to death
F.
G.
H.
Mitotic bodies
Frequency
I.
Nuclear grade
J.
Tubule formation
K.
Histological grade
L.
Tumour infiltrating lymphocytes
p=0.0373
Tumor fraction
M.
Tumor stage
T1
T2
T3
T4

Figure 1

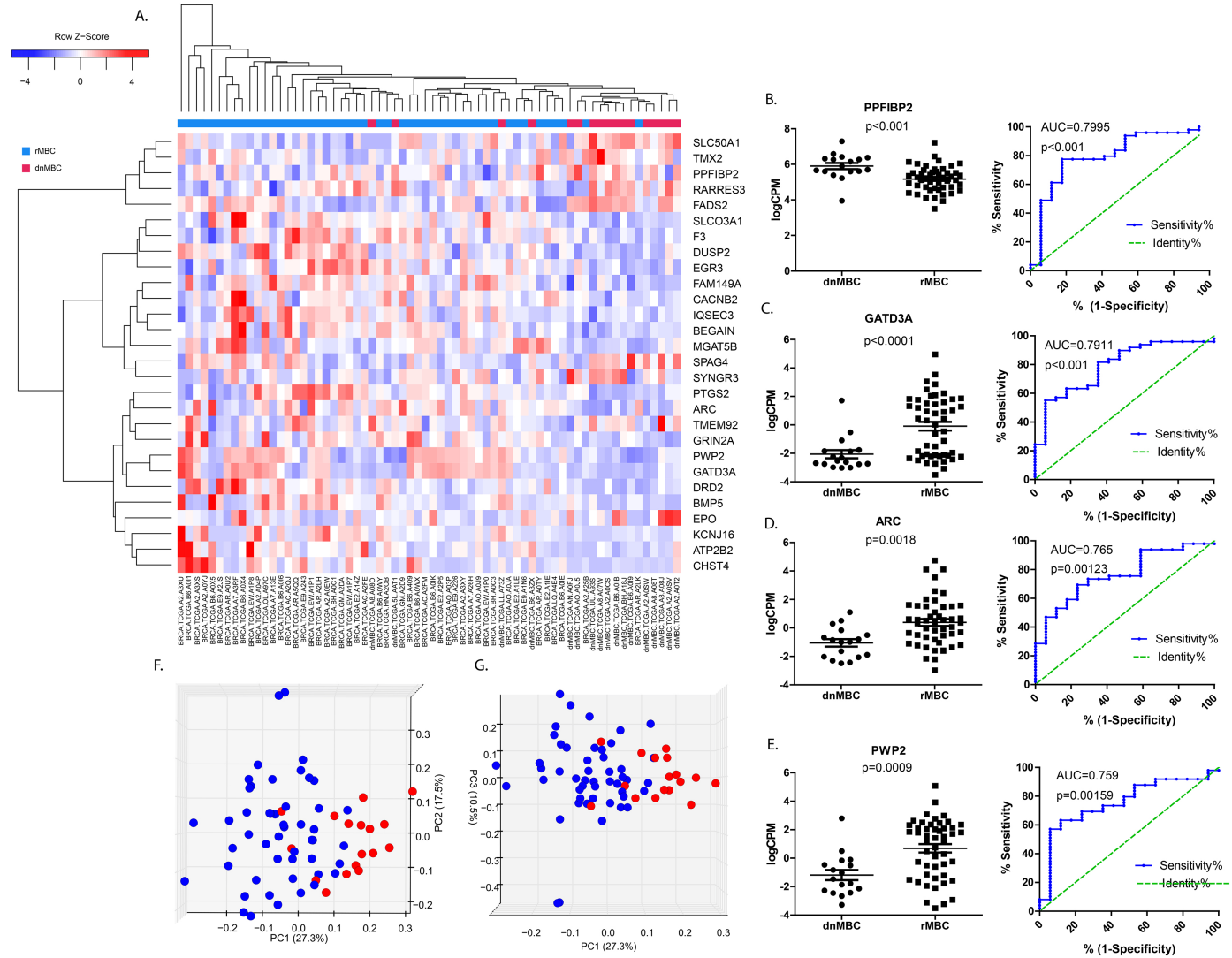
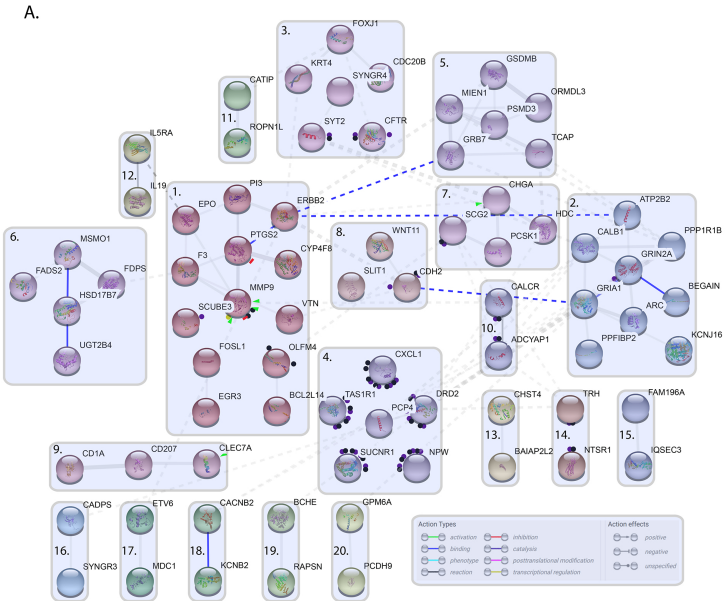
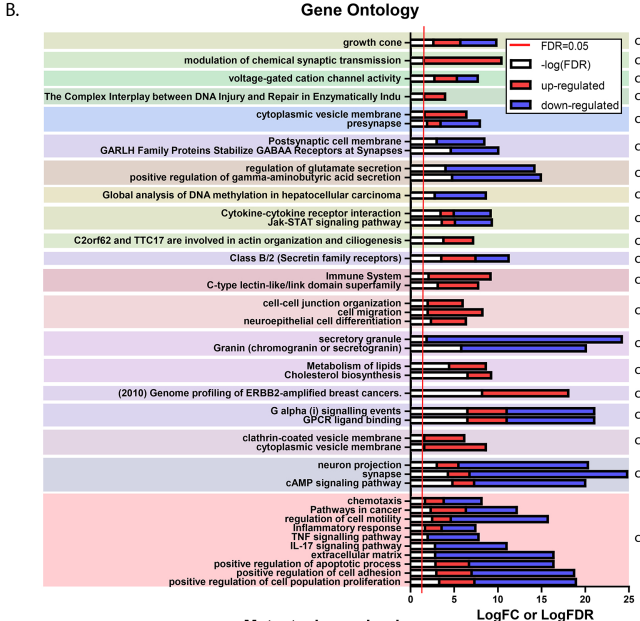
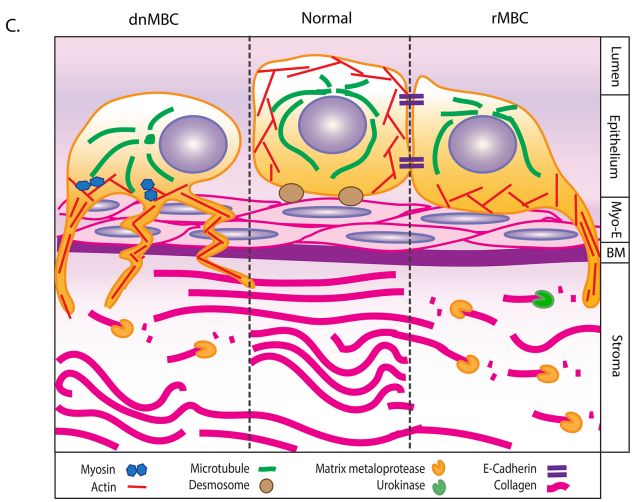

A.
Row Z-Score
−4
0
4
rMBC
dnMBC
SLC50A1
TMX2
PPFIBP2
RARRES3
FADS2
SLCO3A1
F3
DUSP2
EGR3
FAM149A
CACNB2
IQSEC3
BEGAIN
MGAT5B
SPAG4
SYNGR3
PTGS2
ARC
TMEM92
GRIN2A
PWP2
GATD3A
DRD2
BMP5
EPO
KCNJ16
ATP2B2
CHST4
B.
PPFIBP2
p<0.001
logCPM
dnMBC
rMBC
AUC=0.7995
p<0.001
% Sensitivity
% (1-Specificity)
Sensitivity%
Identity%
C.
GATD3A
p<0.0001
AUC=0.7911
p<0.001
D.
ARC
p=0.0018
AUC=0.765
p=0.00123
E.
PWP2
p=0.0009
AUC=0.759
p=0.00159
F.
PC1 (27.3%)
PC2 (17.5%)
G.
PC1 (27.3%)
PC3 (10.5%)

Figure 2


A.
1. PI3 EPO ERBB2 PTGS2 F3 CYP4F8 MMP9 SCUBE3 VTN FOSL1 OLFM4 EGR3 BCL2L14
2. ATP2B2 CALB1 PPP1R1B GRIN2A GRIA1 BEGAIN ARC PPFIBP2 KCNJ16
3. FOXJ1 KRT4 CDC20B SYNGR4 SYT2 CFTR
4. CXCL1 TAS1R1 DRD2 PCP4 SUCNR1 NPW
5. GSDMB MIEN1 ORMDL3 PSMD3 GRB7 TCAP
6. MSMO1 FADS2 FDPS HSD17B7 UGT2B4
7. CHGA SCG2 HDC PCSK1
8. WNT11 SLIT1 CDH2
9. CD1A CD207 CLEC7A
10. CALCR ADCYAP1
11. GATIP ROPN1L
12. IL5RA IL19
13. CHST4 BAIAP2L2
14. TRH NTSR1
15. FAM196A IQSEC3
16. CADPS SYNGR3
17. ETV6 MDC1
18. CACNB2 KCNB2
19. BCHE RAPSN
20. GPM6A PCDH9
Action Types
activation
inhibition
binding
catalysis
phenotype
posttranslational modification
reaction
transcriptional regulation
Action effects
positive
negative
unspecified


B.
Gene Ontology
FDR=0.05
-log(FDR)
up-regulated
down-regulated
growth cone c20
modulation of chemical synaptic transmission c19
voltage-gated cation channel activity c18
The Complex Interplay between DNA Injury and Repair in Enzymatically Indu c17
cytoplasmic vesicle membrane
presynapse c17
Postsynaptic cell membrane
GARLH Family Proteins Stabilize GABAA Receptors at Synapses c16
regulation of glutamate secretion
positive regulation of gamma-aminobutyric acid secretion c15
Global analysis of DNA methylation in hepatocellular carcinoma c14
Cytokine-cytokine receptor interaction
Jak-STAT signaling pathway c12
C2orf62 and TTC17 are involved in actin organization and ciliogenesis c11
Class B/2 (Secretin family receptors) c10
Immune System
C-type lectin-like/link domain superfamily c9
cell-cell junction organization
cell migration
neuroepithelial cell differentiation c8
secretory granule
Granin (chromogranin or secretogranin) c7
Metabolism of lipids
Cholesterol biosynthesis c6
(2010) Genome profiling of ERBB2-amplified breast cancers. c5
G alpha (i) signalling events
GPCR ligand binding c4
clathrin-coated vesicle membrane
cytoplasmic vesicle membrane c3
neuron projection
synapse
cAMP signaling pathway c2
chemotaxis
Pathways in cancer
regulation of cell motility
Inflammatory response
TNF signaling pathway
IL-17 signaling pathway
extracellular matrix
positive regulation of apoptotic process
positive regulation of cell adhesion
positive regulation of cell population proliferation c1
0 5 10 15 20 25
LogFC or LogFDR


C.
dnMBC
Normal
rMBC
Lumen
Epithelium
Myo-E
BM
Stroma
Myosin
Actin
Microtubule
Desmosome
Matrix metaloprotease
Urokinase
E-Cadherin
Collagen

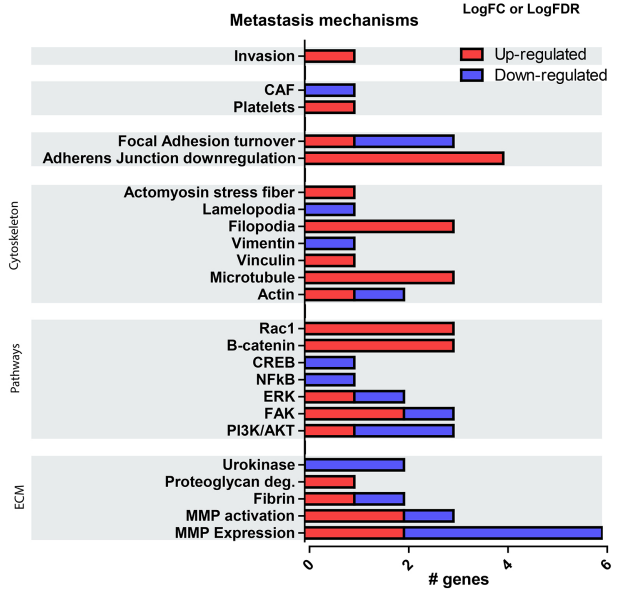

D.
Metastasis mechanisms
Up-regulated
Down-regulated
Invasion
CAF
Platelets
Focal Adhesion turnover
Adherens Junction downregulation
Actomyosin stress fiber
Lamelopodia
Filopodia
Vimentin
Vinculin
Microtubule
Actin
Rac1
B-catenin
CREB
NFkB
ERK
FAK
PI3K/AKT
Urokinase
Proteoglycan deg.
Fibrin
MMP activation
MMP Expression
Cytoskeleton
Pathways
ECM
0 2 4 6
genes

Figure 3

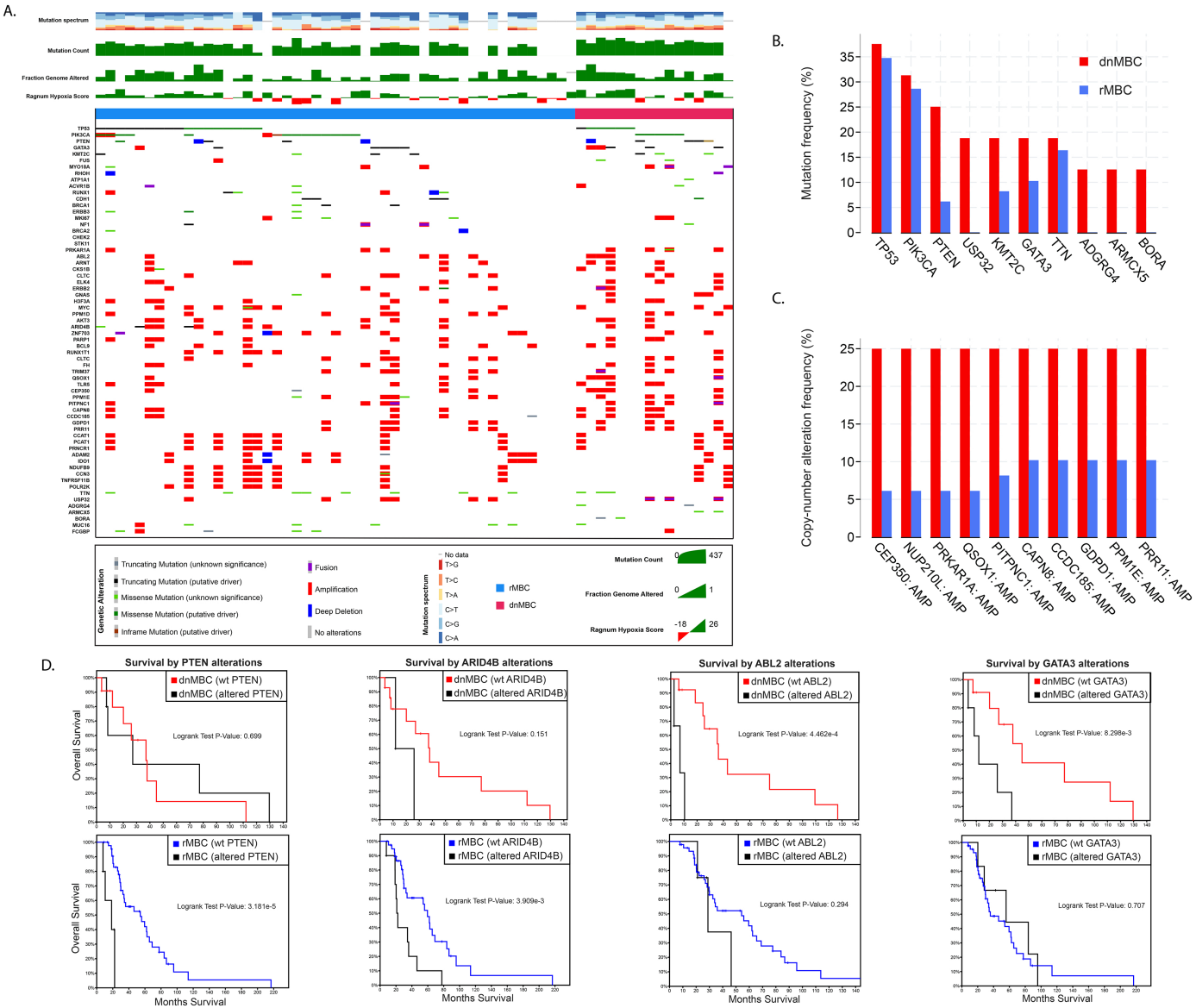

A.
Mutation spectrum
Mutation Count
Fraction Genome Altered
Ragnum Hypoxia Score
TP53
PIK3CA
PTEN
GATA3
KMT2C
FUS
MYO18A
RHOH
ATP1A1
ACVR1B
RUNX1
CDH1
BRCA1
ERBB3
MKI67
NF1
BRCA2
CHEK2
STK11
PRKAR1A
ABL2
ARNT
CKS1B
CLTC
ELK4
ERBB2
CRAC
H3F3A
MYC
PPM1D
AKT3
ARID4B
ZNF703
PARP1
BCL9
RUNX1T1
CLTC
FH
TRIM37
QSOX1
TLR8
CEP350
PPM1E
PITPNC1
CAPN8
CCDC185
GDPD1
PRR11
CCAT1
PCAT1
PRNCR1
ADAM2
IDO1
NDUFB9
CCNS
TNFRSF11B
POLR3K
TTN
USP32
ADGRG4
ARMCX5
BORA
MUC16
FCGBP
Genetic Alteration
Truncating Mutation (unknown significance)
Truncating Mutation (putative driver)
Missense Mutation (unknown significance)
Missense Mutation (putative driver)
Inframe Mutation (putative driver)
Fusion
Amplification
Deep Deletion
No alterations
Mutation spectrum
No data
T>G
T>C
T>A
C>T
C>G
C>A
rMBC
dnMBC
Mutation Count
0
437
Fraction Genome Altered
0
1
Ragnum Hypoxia Score
-18
26
B.
Mutation frequency (%)
dnMBC
rMBC
TP53
PIK3CA
PTEN
USP32
KMT2C
GATA3
TTN
ADGRG4
ARMCX5
BORA
C.
Copy-number alteration frequency (%)
CEP350: AMP
NUP210L: AMP
PRKAR1A: AMP
QSOX1: AMP
PITPNC1: AMP
CAPN8: AMP
CCDC185: AMP
GDPD1: AMP
PPM1E: AMP
PRR11: AMP
D.
Survival by PTEN alterations
dnMBC (wt PTEN)
dnMBC (altered PTEN)
Logrank Test P-Value: 0.699
rMBC (wt PTEN)
rMBC (altered PTEN)
Logrank Test P-Value: 3.181e-5
Overall Survival
Months Survival
Survival by ARID4B alterations
dnMBC (wt ARID4B)
dnMBC (altered ARID4B)
Logrank Test P-Value: 0.151
rMBC (wt ARID4B)
rMBC (altered ARID4B)
Logrank Test P-Value: 3.909e-3
Survival by ABL2 alterations
dnMBC (wt ABL2)
dnMBC (altered ABL2)
Logrank Test P-Value: 4.462e-4
rMBC (wt ABL2)
rMBC (altered ABL2)
Logrank Test P-Value: 0.294
Survival by GATA3 alterations
dnMBC (wt GATA3)
dnMBC (altered GATA3)
Logrank Test P-Value: 8.298e-3
rMBC (wt GATA3)
rMBC (altered GATA3)
Logrank Test P-Value: 0.707

Figure 4

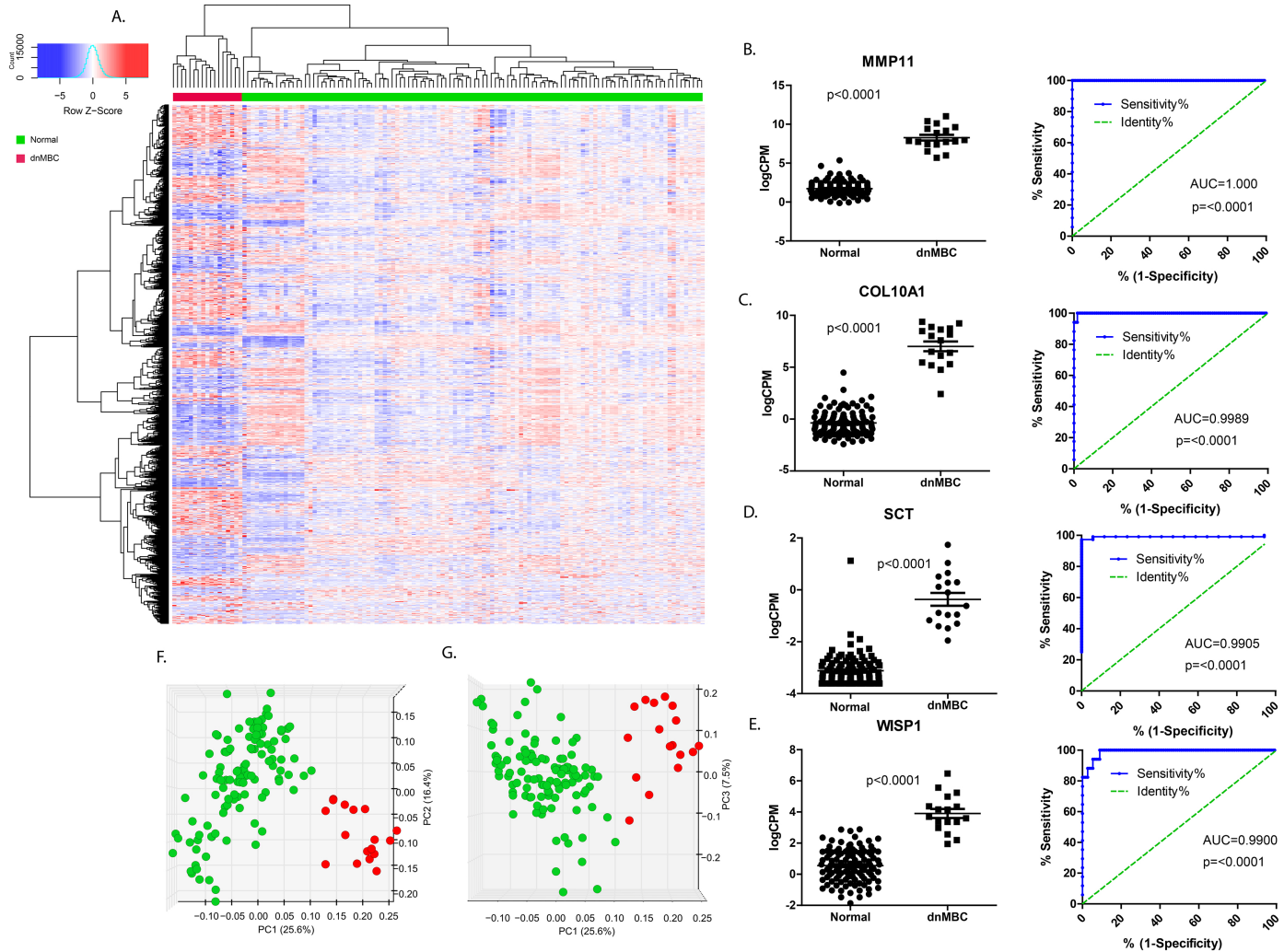

A.
Count
15000
0
−5
0
5
Row Z−Score
Normal
dnMBC
B.
MMP11
p<0.0001
logCPM
Normal
dnMBC
Sensitivity%
Identity%
AUC=1.000
p=<0.0001
% Sensitivity
% (1-Specificity)
C.
COL10A1
p<0.0001
AUC=0.9989
p=<0.0001
D.
SCT
p<0.0001
AUC=0.9905
p=<0.0001
E.
WISP1
p<0.0001
AUC=0.9900
p=<0.0001
F.
PC1 (25.6%)
PC2 (16.4%)
G.
PC1 (25.6%)
PC3 (7.5%)

Figure 5