Article

Phenotypic heterogeneity drives differential disease outcome in Triple Negative Breast Cancer

Archana P. Thankamony 1,2, Reshma Murali 1, Priyanka Chakraborty 3, Nitheesh Karthikeyan 1, Binitha Anu Varghese 1, Vishnu Sunil Jaikumar 1, Mohit Kumar Jolly 3 and Radhika Nair 1*


Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Distributed under a Creative Commons CC BY license.
Abstract: The triple-negative breast cancer (TNBC) subtype is one of the most aggressive forms of breast cancer that has poor clinical outcome and still remains as an unmet clinical challenge. Accumulating evidence suggests that intratumoral heterogeneity or the presence of phenotypically heterogeneous cell populations within a tumor plays a crucial role in chemoresistance, tumor progression and metastasis. Increased understanding of the molecular regulators of intratumoral heterogeneity will enable the development of effective therapeutic strategies in TNBC. We have identified a molecular mediator involved in intratumoral heterogeneity in breast cancer using an unbiased approach. We isolated two heterogeneous tumor cell populations from the 4T1 TNBC tumor model and phenotypic characterization revealed that the cells are distinct in terms of their morphology, proliferation and self-renewal ability in vitro; as well as primary tumor formation and metastatic potential in vivo. Further, RNA sequencing on both cell populations was performed to identify the molecular mediators underlying this heterogeneity. Bioinformatic analysis performed on the differentially expressed genes along with the Kaplan-Meier survival analysis in TNBC patients identified Metastasis associated colon cancer 1 (Macc1) as the top candidate gene mediating the aggressive phenotype. The role of Macc1 in regulating the proliferative phenotype was validated using siRNA mediated gene knockdown. The role of Macc1 in the aggressive cancer cell phenotypes and disease progression is being studied further using a small molecule transcriptional inhibitor of Mac1 in cell line and animal models, thus increasing our understanding of the molecular underpinnings of intratumoral heterogeneity in breast cancer that is critical to the improvement in treatment of women currently living with the highly aggressive TNBC subtype.

Keywords: Intratumoral heterogeneity, Triple negative breast cancer (TNBC), Macc1, Phenotypic heterogeneity

1. Introduction

Breast cancer is a highly heterogeneous disease and displays considerable diversity between patients, also known as intertumoral heterogeneity and within individual tumors known as intratumoral heterogeneity [1, 2]. The diversity in the expression of hormonal (ER/PR) and human epidermal growth factor receptor 2 (HER2) between patients has enabled the classification of breast cancer into distinct subtypes and has been the basis of targeted therapies [1, 2]. The triple-negative breast cancer (TNBC) subtype constitutes 10-20% of all breast cancer cases [3, 4] and is considered as the most devastating subtype due to its aggressive nature [5]. Since the TNBC subtype lacks the expression of hormone receptors (ER, PR) and HER2, it does not respond to hormonal or anti-HER2 therapies [6]. Currently there are no molecular targeted therapies against TNBC and chemotherapy is the mainstay treatment option. However, only 20% of TNBC cases respond to conventional chemotherapy [6] and have lower five-year survival rates when compared to the other subtypes [3].

Intratumoral heterogeneity refers to the presence of distinct cell populations within a tumor with varying tumorigenicity, resistance to therapy and metastatic ability [7, 8].

1 Cancer Research Program, Rajiv Gandhi Centre for Biotechnology, Kerala 695014, India; archanpta@rgcb.res.in (A.P.T.); murali.reshmamurali@gmail.com (R.M.); nitheeshpvk@gmail.com (N.K.); binitha.a.v@gmail.com (B.A.V.); vishnusj@rgcb.res.in (V.S.J.)
2 Manipal Academy of Higher Education (MAHE), Manipal, Karnataka 576104, India
3 Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore 560012, India; priyanka08993@gmail.com (P.C.); mkjolly@iisc.ac.in (M.K.J.)
4 Correspondence: radhikanair@rgcb.res.in; Tel.: +91-471-2781251
Present address: Rajiv Gandhi Centre for Biotechnology, Kerala 695014, India
This heterogeneity arises as a consequence of cell intrinsic factors like genetic/epigenetic alterations and could also be due to cell extrinsic or microenvironmental fluctuations [6]. Since intratumoral heterogeneity and the presence of phenotypically distinct tumor cells within a single tumor is critical to cancer progression and resistance to therapy, understanding the molecular mechanisms underlying the intratumor phenotypic heterogeneity is key for therapeutic targeting of TNBC [9].

The distinct tumor cell populations may cooperate to promote tumor aggressiveness as shown by a study on MDA-MB-231 [10]. Using the UbC-StarTrack system, they isolated clonal cell lines from the MDA-MB-231 cell line. Complementation assays, co-injection and co-growth studies using two clones showed enhanced invasion and migration. Interaction between different cell populations could enhance their fitness and ability to adapt to varying conditions including chemotherapy induced stress. Another recent study by Kuiken et al has shown that distinct clonal subpopulations derived from a human TNBC cell line MDA-MB-468 displayed significant functional heterogeneity and also varied in their growth dynamics under different contexts [11]. These studies pinpoint to the importance of isolating heterogeneous tumor cell subpopulations and unravelling the molecular basis of heterogeneity in order to devise therapeutic intervention strategies for TNBC.

In an interesting study, Xiang et al observed two morphologically distinct tumor cell types following the *in vitro* culture of 4T1 primary tumor and lung metastases [12]. One cell type displayed typical epithelial morphology and other cell type was round and loosely touching surrounding cells [12]. However, these cells were not separately isolated or phenotypically characterized. Additionally, the contribution of these two morphologically distinct tumor cells towards the disease outcome was also not studied thoroughly. In our study, we successfully isolated two distinct tumor cell types from the primary tumor which exhibited different *in vitro* phenotypes as well as *in vivo* disease outcomes. We further employed an unbiased transcriptomic analysis of the two cell populations in order to identify the key mediator of this phenotypic heterogeneity, thus opening up exciting new avenues to explore targeting of aggressive subpopulations of cells within a tumor.

2. Materials and Methods

**Cell culture**

Mouse breast cancer cell line, 4T1, stably expressing green fluorescent protein (4T1-GFP) was obtained from Dr. Alexander Swarbrick (Tumor progression research group, Garvan Institute of Medical Research, Australia) was cultured in RPMI 1640 (11875119, Thermo Fischer Scientific) supplemented with 10% FBS (10270106, Thermo Fischer Scientific), Penicillin (100 Units/mL) /Streptomycin (100µg/mL) (15070063, Thermo Fischer Scientific), 0.25 % (v/v) D+ Glucose (G8644, Sigma Aldrich), 1mM sodium pyruvate and 20mM HEPES. The primary cells were cultured in DMEM/F-12 (11320082, Thermo Fisher Scientific) supplemented with 10% FBS and Penicillin (100 Units/mL) /Streptomycin (100µg/mL). All the cells were grown in 5% CO₂ incubator at 37°C.

**In vivo 4T1-GFP tumor model**

All animal experiments were conducted in accordance with RGCB institutional animal ethics committee. In order to generate 4T1-GFP spontaneous metastasis model, six to eight weeks old BALB/c mice were first anesthetized with Isoflurane. A small incision of 3 cm was cut on the skin along the midline and towards right leg. 7x10⁸ 4T1-GFP cells in 10µL DPBS with salts were injected into the fourth mammary fat pad and the wounds were closed with metallic clips. The mice were monitored daily. After the tumor becomes palpable, the tumor growth was measured using vernier callipers and tumor volumes were calculated \((V = (l \times W^2)/2)\), \(V\)- volume, \(l\)- length, \(W\)-width). For the resection
experiment, the primary tumor was surgically removed when the tumor reached 0.8 cm³ volume. For experimental metastasis assay, 5x10⁵ 4T1A and 4T1NA cells in 100 µL DPBS were injected via tail vein (n=5). At ethical end point (as indicated by reduction in body weight and acute respiratory stress), mice were sacrificed and lung metastases were harvested. For long term survival analysis, mice were injected with 7x10⁵ 4T1A and 4T1NA cells (n=5 and n=9 respectively). The tumor growth over time was monitored and mice was sacrificed as each mice reached their ethical endpoint. Survival analysis was performed using GraphPad Prism software. Samples from primary tumor and lung metastases were collected for immunohistochemical analysis and stored for RNA and protein isolation.

Tumor dissociation and flow cytometric analysis of cell populations

Tumors harvested from the mice were dissociated into single cell suspension for FACs sorting. Tumor dissociation into single cell suspension was carried out using the MACS mouse Tissue Dissociation Kit (Miltenyi Biotec, USA) using manufacturer’s protocols. Following the dissociation approximately 1x10⁶ of tumor cells were taken and washed once with DPBS with salts. The cells were then incubated in FcR blocking reagent (130-092-575, Miltenyi Biotec, USA) for 10 minutes on ice. After incubation, the cells were washed twice using DPBS with salts. Cells were stained for 20 minutes on ice with fluorochrome conjugated antibodies against lineage markers Epcam-FITC (130-102-995, Miltenyi Biotec, USA), CD140a-PE (130-102-502, Miltenyi Biotec, USA). Prior to analysis the cells were washed twice and resuspended in FACs buffer (DPBS with salts + 20% FBS + 20% HEPES). The stained tumor cells were analyzed using the flow cytometer (BD FACS Aria III, BD Biosciences, San Jose, CA, USA).

Cell proliferation assay

5x10² 4T1Round and 4T1Adherent cells were seeded into 96 well plates and the cell proliferation were measured at for 3 days using MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, M6494, Thermo Fischer Scientific) assay. To each well, 10 µl MTT solution (5mg/ml) was added and incubated for 2 hours at 37°C. After incubation, the plates were centrifuged at 1200rpm for 5 minutes at room temperature and the media was removed. The purple-colored formazan crystals were dissolved by adding 100 µl DMSO (D5879, Sigma Aldrich) and the absorbance at 570 nm was measured using plate reader (Tecan).

Tumorsphere assay

For sphere assay using the phenotypically distinct primary cells, 4T1Adherent (4T1A) and 4T1Round (4T1NA) cells were trypsinized and washed with DPBS. These cell populations were then resuspended in serum free DMEMF12 media supplemented with 1X B27 (17504-044, Thermo Fischer Scientific), 20ng/mL bFGF (GF003, Sigma Aldrich), 4µg/mL Heparin (H3149-50KU, Sigma Aldrich). 4T1Round cells were filtered using a 40um cell strainer to avoid cell clumps. Both the cells were seeded at a density of 1x10⁵ cells per well in an ultra-low attachment 24 well plate (Corning, Sigma). Media was replenished every third day and the spheres were counted by day 5.

Quantitative real time PCR (qRT-PCR)

Total RNA was isolated using Trizol® reagent (15596026, Thermo Fischer Scientific) and 1.5µg of RNA was used for cDNA preparation using the High-Capacity cDNA Reverse Transcription Kit (4374966, Thermo fisher Scientific). Real-time PCR was performed using iq™ Umiq SYBR Green (1725121, BioRad) on the QuantStudio 7 Flex RealTime PCR System (Applied Biosystems). The PCR conditions are 95°C for 10 min, followed by 40 cycles of 95°C for 30 s, and 60°C for 1 minute. All reactions were performed in triplicates and
the transcript levels are normalized to those of β-actin. The relative fold change is determined by $2^{-\Delta\Delta CT}$ method.

**Hematoxylin & Eosin (H & E) Staining**

Breast tumor tissues from mice were fixed in 10% neutral buffered formalin (24-72 hours). The fixed tissues were then put into cassettes and transferred into PBS. Samples were embedded in paraffin and 5µm thick sections were taken. Subsequently, the tissue sections were stained using Hematoxylin and Eosin (H & E) and examined for metastases. After incubating at 60°C for one hour, the sections were deparaffinized in xylene followed by rehydration in graded isopropanol solutions. The slides were then rinsed in tap water and kept in acid alcohol for 3 seconds again followed by tap water wash. For the bluing step, the slides were dipped in bluing solution three-four times. Afterwards, the slides were rinsed thrice with tap water and transferred into 70% isopropanol for 3 minutes. Following this, the sections were stained with Eosin Y for 10 minutes. After staining, dehydration was performed in graded isopropanol followed by clearing in xylene. The slides were then mounted with DPX and were visualized using bright field microscope (Leica).

**RNA sequencing**

Total RNA from three replicates each of 4T1GFP (C1, C2, C3), 4T1A (N1, N2, N3), 4T1NA (H1, H2, H3) were isolated using the RNeasy minikit (74104, Qiagen) according to manufacturer’s instructions. RNA concentration of the samples were estimated using Qubit RNA Assay BR (Q10211, Invitrogen). RNA purity was determined using QIAxpert and RNA integrity was measured using RNA ScreenTapes (5067-5576, Agilent) on TapeStation. All the 9 RNA samples passed the QC and were used for the library prep. A modified Illumina TruSeq RNA Sample Preparation V2 protocol (RS1222001) was used for making RNA sequencing libraries. The initial RNA Concentration of 500ng was taken. First, the total RNA sample was subjected to mRNA enrichment by using Poly(A) purification beads. After purification, the mRNA was fragmented under elevated temperatures using divalent cations followed by cDNA synthesis. Indexed adapters were ligated to the cDNA and was then purified and enriched using the following thermal conditions: initial denaturation 98°C for 30sec; 13 cycles of -98°C for 10sec, 60°C for 30sec, 72°C for 30sec; final extension of 72°C for 5mins. PCR products were then purified and checked on TapeStation D1000 DNA tapes (5067-5582, Agilent) for fragment size distribution. Using Qubit High Sensitivity Assay (Q32852, Invitrogen), the prepared libraries were quantified. Before cluster amplification, the libraries were pooled and diluted to the final loading concentration. After completing the cluster generation, the flow cell was loaded on Illumina HiSeqX instrument and 150bp paired end reads were generated.

**Bioinformatic analysis**

The FASTq files obtained after sequencing was checked for quality using FASTQC software. Following the quality check, the adapters were removed from the sequences using Trimmomatic tool. After trimming the adaptors and subsequent quality check, the reads were aligned to the mouse reference genome (mm10) using TopHat2. HTSeq tool was used to obtain the read counts after alignment. The obtained read counts were normalized and differential gene expression analysis was performed using DESeq2. The genes which showed a false discovery rate (FDR) adjusted p value (q value) < 0.01 and Log2FC >2 were considered statistically significant. Heatmap of the differentially expressed genes were constructed using HeatMap tool in the GenePattern software.

**Kaplan Meier Plotter**
The top 50 upregulated genes with q value < 0.01 and Log2FC >2 were further filtered based on their prognostic significance in triple negative breast cancer (TNBC) patients. The association of candidate gene expression with the survival of TNBC patients were analyzed using Kaplan-Meier plotter (KM plotter) database (http://kmplot.com/analysis/). Patient cohorts were split on median gene expression by auto-selecting best cutoff. The association of relapse free survival (RFS) with each candidate genes were determined. The hazard ratio (with confidence intervals) and p values were obtained from the KM plotter. 5 candidate genes were chosen which correlated with significantly (p value < 0.05) lower relapse free survival (RFS) in TNBC patients.

EMT score calculation

The EMT scores of the samples were calculated using 76 gene signature, KS and MLR methods as described previously [13]. The normalized counts of individual genes of all the samples were used for the score calculation. EMT scores for 4T1-GFP, 4T1^A and 4T1^NA cells were calculated using all the three methods.

siRNA mediated knockdown of Macc1

6x10^5 4T1^A cells were seeded in 96 well plate. For small interference RNA (siRNA) mediated knockdown of Macc1, 4T1^A cells were reverse transfected with 50nM of universal negative control (control) and targeting MISSION® Predesigned siRNA (Sigma Aldrich) using Lipofectamine 2000 following manufacturer’s instructions in optiMEM (31985062, Thermo Fisher Scientific) media. After 24 hours, media was changed with fresh DMEM/F-12 media. The cell proliferation 48 hours post transfection was analyzed by MTT assay. For RNA isolation, 2.5x10^5 4T1^A cells were seeded in 6 well plate. 48 hours after transfection, RNA was isolated from the plates using Trizol method.

Statistical Analysis

All experiments were performed in three biological replicates. Statistical analyses were performed using GraphPad Prism 6 (GraphPad Software, Inc., San Diego, CA, USA). Data represented are means ± standard deviation. Statistical tests used are unpaired Student’s t-test and two-way ANOVA. p-values < 0.05 were considered statistically significant with * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.
3. Results

3.1. Two distinct tumor cell populations were obtained from the primary tumor

We used the 4T1 syngeneic TNBC mouse model to isolate heterogeneous tumor cell populations based on previous observations by Xiang et al. [12]. Our strategy used GFP tagged 4T1 cells (4T1GFP), which were orthotopically injected into the 4th mammary fat pad of 6-8 weeks BALB/c mice (Figure S1). We were able to separate the tumor and stromal cells from the primary tumor and metastasis based on the expression of GFP. The primary tumor was surgically removed one month post injection when the tumor reached the ethical volume of 1cm³ (Figure 1 a) to mimic the resection that patients undergo in a clinical setting. Following this, the primary tumor was dissociated and sorted based on the expression of GFP and EPCAM. The GFP -ve stromal population was further sorted based on the expression of CD140a to obtain cancer associated fibroblasts (CAFs) (GFP-/CD140a+). When we cultured these CAFs, we noticed GFP +ve tumor cells in the culture. After three passages, these GFP +ve tumor cells outgrew the CAFs (Figure S2). We then sorted these cells based on GFP. Two tumor cell populations, that showed distinct GFP expression levels were obtained- a GFP-High and a GFP-Low cell populations (GFP negative cell population was not taken as it also contains some CAF cells). Interestingly, these a GFP-High and a GFP-Low tumor cell populations also displayed different morphological appearance when cultured in vitro which was similar to observations made by Xiang and colleagues (Figure 1 b). The GFP-High cells showed a round and loosely adherent morphology whereas GFP-Low cell population exhibited adherent, 4T1-like morphology. Based on their morphology we renamed the two distinct cells as 4T1 Adherent and 4T1 Round.

We were also interested in isolating differentially expressing GFP tumor cells from matched lung metastasis. In order to achieve this aim, the mice were sacrificed and the lungs were harvested and examined for metastasis once ethical end point was indicated by deteriorating body score condition. Like the primary tumor, 4T1 Adherent and 4T1 Round cell populations were also obtained from the matched lung metastasis sample. For the purpose of this work, we will focus on the phenotypically distinct tumor cells obtained from the primary tumor.
Figure 1. Isolation of heterogeneous tumor cell populations from primary tumor. (a) Schematic showing the dissociation and sorting of primary tumor based on the expression of GFP and EPCAM to obtain CAFs. Few GFP+ve tumor cells came along with CAFs after sorting, they were further sorted based on GFP expression. Scatter plot shows the sorting of CAF-tumor cells based on GFP expression using flow cytometry. Two distinct tumor cell populations were obtained following the sorting- and adherent cell population named as 4T1Adherent and a round cell population named as 4T1Round. (b) Representative phase contrast and fluorescent microscopy images showing the 4T1Adherent and 4T1Round tumor cell populations. The primary tumors were dissociated and sorted after resection. The GFP-High and GFP-Low tumor cells obtained after sorting displayed round and adherent morphology respectively when cultured in vitro. (Magnification 10x, 20x)

3.2. *In vitro characterization revealed that 4T1Round and 4T1Adherent tumor cell populations are phenotypically heterogeneous*

To study the intratumoral heterogeneity at a functional level, in vitro phenotypic characterization of the round and adherent tumor cell populations was performed. Proliferation analysis using MTT revealed that the 4T1Adherent showed significantly higher proliferation rate compared to the 4T1Round (Figure 2a). To assess the self-renewal ability tumorsphere assay was performed on the two tumor cell types which are now termed 4T1Adherent and 4T1Round based on their morphology. 4T1Round cells showed significantly higher self-renewal ability compared to 4T1Adherent cells as indicated by the number of 4T1Adherent tumorspheres (Figure 2b). In addition, the 4T1Adherent and 4T1Round tumor cells were also distinct in terms of the expression of epithelial markers (Epcam, Ecadherin), mesenchymal markers (Vimentin, Pdtra) and dormancy associated marker (Nrr2f1) as shown by the qRT-PCR analysis (Figure 2c). The 4T1Adherent cells showed significantly higher expression of epithelial markers whereas the 4T1Round cells showed significantly higher expression of mesenchymal markers. This shows that the 4T1Round tumor cells are more mesenchymal when compared to the 4T1Adherent cells.
Figure 2. Phenotypic characterization of 4T1<sub>Adherent</sub> and 4T1<sub>Round</sub> tumor populations from primary tumor (a) Proliferation analysis using MTT assay shows that 4T1<sub>Adherent</sub> cells proliferated at a significantly faster rate. (b) Quantification of the number of tumorspheres formed by the 4T1<sub>Adherent</sub> and 4T1<sub>Round</sub> populations in tumor-sphere assay. Phase contrast and fluorescent microscopy images showed that 4T1<sub>Adherent</sub> cells formed fewer and smaller spheres (c) Quantitative real-time PCR showing the relative mRNA expression of epithelial-mesenchymal and quiescence genes in the 4T1<sub>Adherent</sub> and 4T1<sub>Round</sub> populations from the primary tumor. Data are expressed as mean ± standard deviation. n=3, p <0.05 are considered statistically significant with *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

3.2. 4T1<sub>Round</sub> and 4T1<sub>Adherent</sub> cells showed distinct disease outcome

The results from the in vitro characterization clearly showed that we had successfully isolated two distinct tumor cell populations. We next addressed if phenotypic heterogeneity translated to different tumorigenic capacity in vivo. The 4T1<sub>Round</sub> and 4T1<sub>Adherent</sub> tumor cells were orthotopically injected into the mammary fat-pad of six-eight week old BALB/c mice and the primary tumor growth was monitored over 30 days. Interestingly all the mice (5/5) that were injected with the 4T1<sub>Adherent</sub> tumor cells formed primary tumors. However, only 2 out of the 9 mice injected with the 4T1<sub>Round</sub> cells formed tumors, that were significantly smaller (Figure 3 a, b). This fits in well with our observation that 4T1<sub>Adherent</sub> tumor cells showed higher proliferation rate in vitro when compared to 4T1<sub>Round</sub> cells. In addition, we performed the Kaplan Meier survival analysis on the mice injected with 4T1<sub>Round</sub> and 4T1<sub>Adherent</sub> tumor cells (Figure 3 c). Mice injected with the 4T1<sub>Adherent</sub> tumor cells showed a median survival of 31 days whereas the mice injected with 4T1<sub>Round</sub> cells showed a median survival of 150 days. Log-rank test revealed that mice injected with 4T1<sub>Adherent</sub> tumor...
cells had significantly lower survival rate compared to mice injected with the 4T1\textsuperscript{Round} tumor cells (p = 0.0004). We also analysed the metastatic potential of these two cell types using the experimental metastasis assay. We injected the 4T1\textsuperscript{Round} and 4T1\textsuperscript{Adherent} tumor cells via the lateral tail vein of BALB/c mice (n=5 each). After sacrificing the mice, lungs were harvested and examined for metastases. As shown by the representative H&E staining, all the mice injected with the 4T1\textsuperscript{Adherent} cells formed large lung macrometastases (5/5) which could not be scored (Figure 3 d). The mice injected with the 4T1\textsuperscript{Round} cells did not exhibit lung metastasis in all samples and similar to the primary tumor result, we found the metastases formed were smaller in size. The in vivo results clearly showed that the 4T1\textsuperscript{Adherent} tumor cells have significantly higher tumorigenicity and metastatic potential compared to the 4T1\textsuperscript{Round} cells. Based on the disease phenotypes observed, we have changed the terminology for the 4T1\textsuperscript{Adherent} tumor cells to 4T1\textsuperscript{A} (aggressive) and 4T1\textsuperscript{Round} cells to 4T1\textsuperscript{NA} (non-aggressive) for ease of association of the phenotype of two cell types with their genotype in the subsequent work. The in vitro and in vivo phenotypes of the 4T1\textsuperscript{A} and 4T1\textsuperscript{NA} cells are summarized in Table S1.

**Figure 3.** 4T1\textsuperscript{Adherent} cells displayed highly aggressive disease outcome a) In vivo tumorigenic potential of 4T1\textsuperscript{Adherent} and 4T1\textsuperscript{Round} tumor cells were analyzed by injecting 7x10\textsuperscript{6} tumor cells orthotopically into the fourth mammary fat-pad of BALB/c mice. Table shows the number of mice that developed the primary tumors. Representative image shows the primary tumors formed by 4T1\textsuperscript{Adherent} and 4T1\textsuperscript{Round} tumor cells after 4 weeks. The 4T1\textsuperscript{Adherent} cells are highly aggressive as they formed larger primary tumors compared to the 4T1\textsuperscript{Round} cells (b) Tumor growth curves of mice injected with 4T1\textsuperscript{Adherent} (n=5) and 4T1\textsuperscript{Round} cells (n=9) (c) Kaplan-Meier survival curves of mice injected with 4T1\textsuperscript{Adherent} and 4T1\textsuperscript{Round} tumor cells. (d) Representative hematoxylin and eosin staining of lungs of mice injected with 5x10\textsuperscript{5} 4T1\textsuperscript{Adherent} and 4T1\textsuperscript{Round} tumor cells via the lateral tail veins (n=5). The black dashed lines and arrows indicating the metastatic lesions. Magnification 10x.
Global transcriptomic analysis of the 4T1A and 4T1NA cells identifies molecular mediators of phenotypic heterogeneity

To identify the differentially expressed genes involved in mediating the phenotypic heterogeneity in the 4T1 primary tumor and aggressive phenotype of 4T1A cells compared to the 4T1NA cells, we performed transcriptomic analysis on the 4T1GFP, 4T1A and 4T1NA cells using RNA sequencing. RNA sequencing was performed using illumina HiSeqX platform. To further analyse the raw sequencing data, we employed widely used software tools such as TopHat2, HTSeq-count and DESeq2 from the Galaxy platform (www.usegalaxy.org) using default settings. The read quality was analyzed using FASTQC. Following this, the reads were filtered and processed using Trimmomatic tool which removes the adaptors and low-quality reads. The filtered reads were aligned to mouse reference genome mm10 using TopHat2. The read counts were obtained from HTSeq Count tool. The read counts were normalized and differential gene expression analysis was performed using DESeq2 which uses median of ratios as normalization method. The top candidate genes were then identified based on the filters such as q-value < 0.01 and Log2 Fold change >2 (Figure S3). Based on the filters top 50 candidate genes upregulated in the 4T1A compared 4T1NA population were obtained (Table S2).

EMT score analysis of global gene expression data reveals the mesenchymal nature of 4T1NA cells

We analyzed the EMT status of the three cell populations by applying three different scoring methods – a 76 gene EMT signature (76GS), Kolmogorov Smirnov (KS) method and Multinomial logistic regression (MLR) method [13] on the gene expression data, EMT score was calculated which gives a quantitative measurement of the epithelial-mesenchymal status of the samples. According to the 76 GS method, the absolute EMT score of epithelial (E) sample will be relatively higher than that of mesenchymal (M) sample. In KS method, the EMT score varies from -1 to 1. Higher EMT score in KS method corresponds to more M samples. MLR method identifies the EMT status of a cell based on a scale from 0-2 in which higher score indicates that the sample is more mesenchymal. Interestingly, the EMT score calculation by all the three methods indicates that the 4T1NA cell population was as found to be more mesenchymal than both 4T1GFP parental cell line and 4T1A cells (Figure 5) in consensus with the invitro result.

Figure 4. 4T1NA cells are more mesenchymal Bar plots showing the EMT scores calculated for 4T1GFP, 4T1A, 4T1NA samples using three transcriptomics based EMT scoring methods (p < 0.05).
Table 1. EMT scores of 4T1GFP, 4T1A, 4T1NA samples

<table>
<thead>
<tr>
<th>Sample names</th>
<th>Sample</th>
<th>76GS</th>
<th>MLR</th>
<th>KS</th>
</tr>
</thead>
<tbody>
<tr>
<td>N3</td>
<td></td>
<td>24.8464719</td>
<td>1.0381426</td>
<td>0.1736343</td>
</tr>
<tr>
<td>N2</td>
<td>4T1A</td>
<td>21.9802142</td>
<td>1.0704948</td>
<td>0.190078</td>
</tr>
<tr>
<td>N1</td>
<td></td>
<td>21.0816129</td>
<td>1.062039</td>
<td>0.1683389</td>
</tr>
<tr>
<td>H1</td>
<td></td>
<td>-37.450694</td>
<td>1.2640291</td>
<td>0.2934783</td>
</tr>
<tr>
<td>H2</td>
<td>4T1NA</td>
<td>-36.86977</td>
<td>1.2184623</td>
<td>0.3062988</td>
</tr>
<tr>
<td>H3</td>
<td></td>
<td>-37.275363</td>
<td>1.2664043</td>
<td>0.2845596</td>
</tr>
<tr>
<td>C1</td>
<td></td>
<td>11.9937033</td>
<td>1.0375892</td>
<td>0.201783</td>
</tr>
<tr>
<td>C3</td>
<td>4T1GFP</td>
<td>17.1188572</td>
<td>0.9607885</td>
<td>0.1939799</td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td>14.5749679</td>
<td>1.0378739</td>
<td>0.1953735</td>
</tr>
</tbody>
</table>

Identification of candidate genes associated with survival of TNBC patients

In order to identify the top candidate genes contributing to the aggressive disease phenotype in the 4T1A cell population, we analyzed the association of candidate gene expression with relapse-free survival (RFS) in Triple Negative Breast Cancer (TNBC) patients. RFS of the top 50 candidate genes (Table S2) were analyzed using Kaplan Meier (KM plotter) software. A list of 5 genes which showed a significant association with poor survival in TNBC patients were obtained (Table S3). We selected Macc1 as the top gene (relapse-free survival - Hazards ratio (HR): 2.15, log p value: 0.028) for further validation (Figure 12 a) as TNBC patients with a higher mean expression of MACC1 showed poor relapse-free and distant metastasis-free survival (DMFS) compared to the patients with a lower expression of these genes (Figure 5 a, b).

![Figure 5. Association of relapse and distant metastasis-free survival of TNBC patients with expression of MACC1](image-url)

Correlation of MACC1 gene expression with TNBC patient survival was analyzed using KM plotter. TNBC patients with higher expression (red line) of MACC1 was associated with significant lower relapse-free survival (a) and distant metastasis-free survival (b) compared to the patients with lower expression of MACC1 (black line).
Macc1 depletion caused significant reduction in breast cancer cell proliferation

Metastasis Associated Colon Cancer 1 (MACC1) gene had been initially identified as a novel gene involved in colon cancer progression through HGF/c-MET signaling and predicted poor prognosis [14]. Increased MACC1 expression was later found to be associated with poor prognosis in multiple cancers including breast cancer [15-18]. MACC1 acts as a transcriptional regulator of the protooncogene c-MET and enables the acquisition of a malignant phenotype by promoting the cancer hallmark capabilities through various downstream target molecules [19]. Although the mechanism by which it mediates proliferation, stemness and metastasis is well explored in colorectal and gastric cancer, it still needs to be investigated in breast cancer.

To investigate the role of Macc1 gene expression in breast cancer cell proliferation, we performed siRNA-based silencing of Macc1 in the aggressive 4T1A cells. To validate siRNA mediated knockdown, we analyzed the expression of Macc1 in the knockdown samples using qRT-PCR. A significant reduction in the expression of Macc1 was observed 48hrs post-transfection (Figure 6 a). Interestingly, Macc1 silencing resulted in a significantly lower cell viability (Figure 6 b) in the aggressive tumor cells compared to control. We will further extend this study in a therapeutic context by using Lovastatin, a small molecule transcriptional inhibitor of Macc1 [20, 21] in cell line and mice model in the future (Figure 6 d).

4. Discussion

Breast cancer manifests itself in different clinico-pathological forms depending on underlying gene signatures. Inter tumoral heterogeneity has been at the heart of treatment
modalities of breast cancer with the success in breast cancer management arising largely from the use of targeted therapy like Tamoxifen in ER+ and Herceptin for the Her2 subtypes. But the increasing spotlight on the different cell types within a single tumor termed intra-tumoral heterogeneity results in the challenging situation where drug-resistant cells escape and lead to relapse of the tumor. This is particularly true in the TNBC molecular subtype which though classified as a single entity based on immunohistochemistry (IHC), actually is a group of diseases. More recent work has also shown that there could be unique “ecotypes” within tumors involving the interaction of cancer cells with the microenvironment [22]. Our work aimed to deconvolute the different cell populations within a single TNBC tumor and try to understand if we could exploit the underlying molecular circuitry for therapeutic purposes.

The 4T1 model had previously been reported to have morphologically distinct cell types which could not be further isolated and studied. We decided to use a strategy of tagging the heterogeneous cell population with a GFP tag for further isolation [23]. Serendipitously, we observed two cell populations among others (data not shown) which had differential GFP expression allowing for easy isolating of different cell types from a single tumor. Once cultured, they also exhibited marked difference in morphology which also translated into differing growth kinetics, self-renewal capacity and EMT molecular markers. Interestingly, this also translated into contrasting disease outcomes in vivo. This confirmed our hypothesis that a tumor is composed of non-identical cells which coexist to give different functionality to the cancer.

Unpublished work from our lab showed that the round 4T1<sup>NA</sup> cells eventually acquired elongated mesenchymal like morphology. As indicated by the EMT score analysis, the round cells are more mesenchymal and probably reaching towards the end of the EMT spectrum, finally achieving elongated mesenchymal like morphology in culture indicating cellular plasticity [24]. Another possibility is that, since even the established cell lines exhibit heterogeneity within themselves, they are just another cell type within the 4T1GFP cell line that we were able to isolate serendipitously from the primary tumor.

Our work also underscores the critical need to target both pools of cells in order to get better total therapeutic response. As proof of principle, we were able to identify a potential therapeutic approach to targeting the aggressive tumor cells by pinpointing Macc1 as a key gene overexpressed in these cells. We are currently expanding this finding to target the cells that are responsible for rapid tumor disease progression with Lovastatin, which is a small molecule transcriptional inhibitor of Macc1.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: Schematic showing the 4T1GFP tumor model; Figure S2: Phenotypic characterization of the heterogeneous tumor cell populations; Table S2: 50 candidate genes upregulated in the 4T1<sup>NA</sup> compared to 4T1<sup>NA</sup> cells; Table S3: Top 5 candidate genes upregulated in the 4T1<sup>NA</sup> associated with poor survival in TNBC patients

Author Contributions: Conceptualization, R.N. and A.P.T.; methodology, A.P.T., R.M., P.C., N.K., B.A.V., V.S.J., M.K.J. and R.N.; validation, A.P.T.; formal analysis, A.P.T., R.M., P.C., N.K., B.A.V., V.S.J.; investigation, R.N.; resources, R.N.; writing—original draft preparation, A.P.T. and R.N.; writing—review and editing, all authors; visualization, A.P.T., N.K.; supervision, R.N.; project administration, R.N.; funding acquisition, R.N. All authors have read and agreed to the published version of the manuscript.

Funding: R.N. is the recipient of the Ramanujan Fellowship from the Government of India (SERB) (SB/S2/RJN/182/2014). M.K.J. was supported by Ramanujan Fellowship (SB/S2/RJN-049/2018) provided by SERB, DST, Government of India, and by InfoSys Foundation, Bangalore. A.P.T. was supported by CSIR-Junior research fellowship.
Institutional Review Board Statement: This study was conducted in accordance with the Institutional Animal Ethics Committee (IAEC) of Rajiv Gandhi Centre for Biotechnology (IAEC/706/RAD/2018).

Informed Consent Statement
Not applicable

Data Availability Statement.
The datasets generated for this study can be found in the GEO Database GSE195995, GSM5857692, GSM5857693, GSM5857694, GSM5857695, GSM5857696 and GSM5857697.

Acknowledgments: We are grateful to Dr. Alexander Swarbrick, Garvan Institute of Medical Research, Australia for 4T1GFP cell line. We thank Surabhi S.V. and Tilak Prasad for helping with flow cytometry, Nanditha C.K. for helping with the histopathology, Aswathy Mary Paul and Bijesh George for helping with the computational analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References


