Article

Multi-Omics Characterization of the Spontaneous Mesenchymal–Epithelial Transition in the PMC42 Breast Cancer Cell Lines

Sugandha Bhatia [1,2,3,*], James Monkman [1,2,3], Tony Blick [1,2,3], Pascal HG Duijf [1,3,4], Shivashankar H. Nagaraj [1,2,3] and Erik W. Thompson [1,2,3,*]

[1] Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD 4059, Australia
[2] School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD 4000, Australia
[3] Translational Research Institute, Brisbane, QLD 4102, Australia
[4] University of Queensland Diamantina Institute, The University of Queensland, Woolloongabba, QLD 4102, Australia
[*] Correspondence: sugandha.bhatia@hdr.qut.edu.au (S.B.); e2.thompson@qut.edu.au (E.W.T.)

check for updates

Abstract: Epithelial–mesenchymal plasticity (EMP), encompassing epithelial–mesenchymal transition (EMT) and mesenchymal–epithelial transition (MET), are considered critical events for cancer metastasis. We investigated chromosomal heterogeneity and chromosomal instability (CIN) profiles of two sister PMC42 breast cancer (BC) cell lines to assess the relationship between their karyotypes and EMP phenotypic plasticity. Karyotyping by GTG banding and exome sequencing were aligned with SWATH quantitative proteomics and existing RNA-sequencing data from the two PMC42 cell lines; the mesenchymal, parental PMC42-ET cell line and the spontaneously epithelially shifted PMC42-LA daughter cell line. These morphologically distinct PMC42 cell lines were also compared with five other BC cell lines (MDA-MB-231, SUM-159, T47D, MCF-7 and MDA-MB-468) for their expression of EMP and cell surface markers, and stemness and metabolic profiles. The findings suggest that the epithelially shifted cell line has a significantly altered ploidy of chromosomes 3 and 13, which is reflected in their transcriptomic and proteomic expression profiles. Loss of the TGFβR2 gene from chromosome 3 in the epithelial daughter cell line inhibits its EMT induction by TGF-β stimulus. Thus, integrative 'omics' characterization established that the PMC42 system is a relevant MET model and provides insights into the regulation of phenotypic plasticity in breast cancer.

Keywords: copy number variations (CNV); epithelial–mesenchymal transition (EMT); karyotyping; mesenchymal–epithelial transition (MET); metabolism; proteomics; RNA-sequencing; seahorse extracellular flux analyser; whole exome sequencing

1. Introduction

Breast cancer, a leading cause of cancer death in women, is recognized as a molecularly heterogeneous disease [1,2]. Breast cancer cell culture systems and patient-derived xenografts (PDX) recapitulate many of the intrinsic molecular subtypes of breast cancer [3]. Various experimental techniques have been employed in these models, such as exome sequencing, copy-number analysis, whole-transcriptome, epigenome and methylome analyses, and identification of biomarkers that were mapped simultaneously to the genotypic and/or phenotypic behaviour in relevance to breast cancer subtypes or mammary gland development [4–15]. The compendium of molecular profiles defining up to 90 different breast cancer cell lines provides a valuable resource and has been studied

extensively [10–13,15], not only for the discovery of new breast cancer genes [10], but also for investigations of their subtype-specific pathobiology, cancer stem cell biology, biomarkers, and response to different drug therapies [16]. Thus, cell culture systems provide important models for studying the molecular mechanisms of neoplastic transformation and are extremely useful in translational research.

Epithelial–mesenchymal plasticity (EMP), which encompasses epithelial–mesenchymal transition (EMT) and its reversal (mesenchymal–epithelial transition; MET), including hybrid states within this spectrum, is considered an important hallmark of cancer that drives metastasis [17–19]. EMT provides carcinoma cells with the ability to undergo cellular morphogenesis, allowing clusters of epithelial cells to form independent motile mesenchymal-like cells that can disseminate. MET is thought to subsequently reinstate proliferation and allow collective outgrowth at the metastatic site. The genetic and phenotypic heterogeneity of carcinoma cells across the EMP axis also has importance in therapy-resistance seen after EMT [20–23]. The molecular and cellular analyses of EMT features in breast cancer cell lines have also been analysed thoroughly and scoring tools have been developed to quantify the extent of EMT [24–29].

EMP studies in cancer cell lines have been widely reported. Therein, EMP is either induced by external stimuli, such as hypoxia [30], TGF-β [31], EGF [32], FGF [33], or by inducible expression or repression of EMT-inducing transcription factors [34,35]. Transformed breast cancer cells obtained by introducing oncogenes or cancer-associated genes into normal primary human mammary epithelial cells, such as HMLE and D492 cells, have also been remodelled as HMLER and D492M, respectively, to study cancer development including EMT [36–39]. To our knowledge, the PMC42 system is the only breast cancer cell line model where a spontaneous MET event has led to a new stable variant [40]. In this regard, the PMC42 system provides a unique case study to investigate the molecular and phenotypic characteristics of a stable MET in carcinoma without the use of external stimuli or genetic manipulation.

The relationships between different levels of regulation (genomic, transcriptomic, metabolomic, etc.) in the intrinsic endogenous plasticity (i.e., not exogenously induced) seen in several EMP systems has not been systematically studied, although each has separately been implicated in EMP regulation. Here, we interrogate and integrate the molecular portrait of PMC42, a unique human breast carcinoma cell line originally established from the pleural effusion of a breast cancer patient in 1983 [41]. Original cultures of PMC42 cells were reported to be heterogeneous, with at least eight different morphological types identified by phase contrast and electron microscopy [41,42]. In earlier studies, PMC42 cells were shown to grow both as a monolayer and as cords in suspension, and have shown features of myoepithelial cells [41,43,44]. PMC42 cells obtained from Dr. Robert Whitehead [41,42] were subsequently annotated as PMC42-ET. PMC42-ET exhibits a 'Basal B' (mesenchymal) transcriptome (E Tomaskovic-Crook and T Blick, unpublished observation) [30] and PMC42-LA cells were a spontaneous derivative cell line from PMC42-ET developed by Dr. Leigh Ackland [40,45]. Although epithelially shifted to the extent that they can produce a functional epithelium in 3-dimensional (3D) cultures [45], they still cluster transcriptomically with PMC42-ET cells (E Tomaskovic-Crook and T Blick, unpublished observation) [30]. This makes these cell lines an ideal system to study MET. Comparison of the karyotype, exome, transcriptome, proteome and metabolic phenotype of PMC42-ET and its epithelial "daughter" cell line PMC42-LA were performed to gain insights into the dynamic events that contributed to this spontaneous MET, in reference to EMP studies.

2. Materials and Methods

2.1. Cell Lines and Cell Culture

Cells (BT-549, Hs578T, MCF-7, T47D, MDA-MB-157, MDA-MB-231, MDA-MB-468, SUM-159) were obtained from the American Type Culture Collection (Manassas, VA, USA). PMC42-ET and the derivative PMC42-LA cell line were derived from a breast cancer pleural effusion with

appropriate institutional ethics clearance (Institutional Review Board of the Peter MacCallum Hospital, Melbourne) [41–43,45].

Cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) containing glucose (4.5 g/L), L-Glutamine (0.5 g/L) and sodium pyruvate (0.1 g/L) (Corning, Catalog number 10-013-CVR), and supplemented with 10% foetal bovine serum (FBS; Gibco^{TM}, Thermo Fischer Scientific, Waltham, MA, USA) and antibiotics penicillin and streptomycin (Gibco^{TM}, Life Technologies, NY, USA; Catalog number 15140122), and maintained at 37 °C, 5% CO_2, and 95% humidity. Cells were routinely tested and were negative for *Mycoplasma*.

2.2. Preparation of Metaphase Spread and Karyotyping

After 60%–70% of confluency was achieved in 60 mm dishes, cells were treated with 10 μL of demecolcine (10 μg/mL) for 3–4 h. Cells were harvested using trypsin (Corning™ 25053CI) and the pellet was gently treated with hypotonic solution (75 mM KCl) for 40–60 min at 37 °C and fixed in cold methanol/acetic acid (3:1). Two or three drops of suspended cells were applied to glass slides and chromosomes were stained with DAPI and imaged by confocal microscopy (Olympus Fluoview FV1200 Confocal Laser Scanning Microscope, Olympus Australia Pty. Ltd., Melbourne, Victoria, Australia). Counting was performed manually in ImageJ. Karyotype assessment via G-band analysis was performed by a commercial genotyping service (StemCore Facility, Brisbane, QLD, Australia) and 50 metaphases were analysed per cell line.

2.3. DNA Extraction, Whole Exome Sequencing and Processing of Sequencing Data

Genomic DNA was extracted from cells using Bioline Isolate II Genomic DNA kit (Cat: BIO-52067) as per the manufacturer's instructions. After quantifying the DNA and checking the purity, DNA samples were shipped to GENEWIZ (Suzhou, China) for whole exome sequencing (WES) and subsequent analysis.

Genewiz, Inc. (Suzhou, China) performed initial quality control assessments and subsequent exome capture using the SureSelectXT HS Target enrichment kit (Agilent). All samples were paired-end multiplex sequenced (2 × 150 base pairs) on the Illumina Hiseq 2500 platform to a median target depth of over 50× WES data have been deposited at NCBI under BioProject ID PRJNA557326.

Paired-end reads underwent quality control before alignment to the reference human genome (hg19) using Burrows-Wheeler alignment (BWA, version 0.7.12-r1039) [46] and SAMtools (version 1.6) [47]. Realignment and recalibration were performed using the Genome Analysis Toolkit (GATK, version 3.5) [48]. Single nucleotide variants (SNVs) and insertions and deletions (indels) were called using GATK with default settings. Annotation of variants (SNP and indels) was performed using ANNOVAR [49,50]. Control-FREEC v 10.6 was used for detecting and filtering the CNV [51].

CRAVAT (Cancer-Related Analysis of Variants Toolkit), a tool specifically tailored to analyze cancer-specific variants [52], was used for identification and prioritization of genes with a possible role in cancer tumorigenesis in the PMC42 cell lines. Identification and annotation of cancer-specific driver missense mutations was performed using CHASM (Cancer-specific High-throughput Annotation of Somatic Mutations) [53,54]. To identify and prioritize pathogenic missense mutations, VEST (Variant Effect Scoring Tool), a supervised machine learning-based classifier [55], was also applied. Both CHASM and VEST computational scores are integrated in the CRAVAT suite.

2.4. RNA Extraction, cDNA Synthesis and RT-qPCR

Total RNA was extracted from cells using TRIzol (Life Technologies) and subsequent reactions were carried out as per the Bioline Isolate II RNA Micro kit manufacturer's instructions, and cDNA was synthesized using SensiFAST^{TM} cDNA Synthesis kit from Bioline. Real time-quantitative PCR (RT-qPCR) was performed using the SYBR Green Master Mix in a ViiA7 Real-Time PCR system (Applied Biosystems) and analysis performed using the Quantstudio^{TM} Real-Time PCR software v1.1 (Applied Biosystems, Life Technologies). The primers sequences are listed in Supplementary Table S1.

Hierarchical clustering of transcriptional profiles was performed using the Morpheus online tool [56]. Data were normalized against the overall mean expression of all measured genes [57].

2.5. Whole Transcriptome Sequencing and Analysis of PMC42 Cell Lines

mRNA transcript abundances for PMC42 cell lines were measured using RNA-seq as previously reported [30]; data have been deposited at NCBI under bioproject ID PRJNA322427. Sequence alignment to the hg19 reference genome was performed using TopHat [58] with default parameters. Differential expression analysis was performed using CuffDiff [58] with default parameters. RNA sequencing results were re-analysed for interrogation of fold changes across the two PMC42 cell lines with respect to the chromosome number. Gene Set Enrichment Analysis [59] was also applied to identify enrichment of gene signatures contained in the Molecular Signatures Database (MSigDB). The filtered gene lists with $p < 0.01$ were also examined by Ingenuity Pathway Analysis®(IPA) for functional annotation and gene network analysis. The GSVA method from the GSVA R/Bioconductor package was also applied on the gene expression data for the PMC42-ET and PMC42-LA cell lines to score samples against the TGFβ-EMT signature.

2.6. Data-Independent Acquisition (DIA) Mass Spectrometry of PMC42 Cell Lines

Cells were washed with ice-cold phosphate buffered saline (PBS), and lysed directly in cell lysis buffer containing 4% (w/v) SDS, 10 mM dithiothreitol (DTT), 10 mM Tris-HCl along with Roche compete protease and phosphatase inhibitors (Roche, Rotkreuz, Switzerland). Lysates were sonicated to shear DNA, and protein concentration was quantified using the Pierce™ BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA). On the basis of protein quantifications, each experimental sample was aliquoted into 25 μg samples for processing using the FASP method [60]. Digestion was performed overnight using Trypsin/Lys-C (Promega) mix in 1:50 of protein. Fragmented peptides were then dissolved in 0.1% formic acid and processed for a final clean-up step using C18 Zip-Tips (Millipore; Billerica, MA, USA).

Protein Pilot (V 4.1) software from SCIEX was used for peptide identification. The human protein library was built using the UniProt database (release 2018_05, [61]) with the following settings: Sample Type, identification; Cysteine alkylation, acrylamide; Instrument, TripleT of 5600; Species, human; ID focus, Biological modification; Enzyme, trypsin; Search effort, thorough ID. False discovery rate (FDR) was calculated within ProteinPilot software and peptides identified with greater than 99% and a local FDR of 1% was applied for the peptide identification. PeakView Software was employed to measure the peptide abundance with standard parameters [62] and manual inspection was carried out to confirm the accuracy of the spectra. Six peptides per protein were used to measure the protein abundance. The differences in protein abundance between PMC42-ET and -LA were calculated based on the significance and fold-changes. MSstats was used to calculate protein level significance by applying a linear mixed-effects model [63]. The model combines quantitative measures for a targeted protein across peptides, charge states, transitions, samples, and conditions; the system detects proteins that change in abundance among conditions more systematically than would be expected by random chance, while controlling the FDR. In house scripts in Python and R were developed for further analysis.

2.7. Fluorescence Activated Cell Sorting (FACS)

Cells were lifted with Accutase®(Corning, Catalog # 25-058-CI) and stained with anti-human CD44-FITC (BD Pharmingen) and anti-human CD24-PB (Exbio) antibodies at manufacturer's recommended dilutions in 0.1% BSA (Bovine serum albumin, Sigma) diluted in DPBS for 1 h in a rotary shaker at room temperature. Cells were analysed in the presence of propidium iodide (1 μg/mL) using a BD LSRFortessa (BD Biosciences). After doublet discrimination and compensation for spectral overlap, data were analysed by using FlowJo Software (BD Biosciences). For TGFβR2 surface expression, cells were stained with primary antibody (RandD Systems, Cat# AF-241-NA) as per manufacture recommended dilutions for 1 h and then with secondary goat antibody for 1 h.

2.8. Immunocytochemistry

The cell lines were seeded at a density of 10,000 cells/well in 48-well plates (Thermo Scientific Nunclon^{TM} Delta Surface-150687). During immunocytochemistry, the growth medium was discarded, and cells were washed thrice gently with Dulbecco's modified PBS (DPBS; pH 7.5). Briefly, cells were fixed in 4% paraformaldehyde ± 0.1% Triton X-100 (depending on the desired permeabilization conditions), rinsed with DPBS, and incubated with primary antibodies at 4 °C overnight. After rinsing again in DPBS, cells were incubated with an appropriate fluorescence-conjugated secondary antibody (Supplementary Table S2) and with diamidino phenyl indole (DAPI) as a nuclear stain (diluted to a final concentration of 1 µg/mL) for 2 h at room temperature in the dark with gentle rotary shaking. The plates were then washed thrice with DPBS and images were captured on a high-content imaging platform (Cytell Cell Imaging System (GE Healthcare) or IN Cell Analyser 6000 (GE Healthcare, Buckinghamshire, UK), as indicated), with approximately 6–9 fields of view taken per well. Images were further analysed and quantified using the IN Cell Investigator software v1.0 (GE Healthcare).

2.9. Seahorse Metabolic Analyser

Collagen (Rat tail, type 1, Cat 354236, BD Biosciences)-coated Seahorse cell culture plates (Seahorse *Bioscience*, 102601-100) were seeded at a density of 20,000 cells per well (XFe96 cell culture microplate; Seahorse Biosciences, North Billerica, MA, USA). The cells were allowed to grow for 24–48 h at 37 °C in 5% CO_2, after which the cells were washed and replaced with assay media (unbuffered DMEM or RPMI supplemented with 10 mM glucose, 1 mM sodium pyruvate, 2mM L-glutamine, no sodium bicarbonate at pH 7.4). The cells were incubated for 1 h at 37 °C in a non-CO_2 incubator. Mitochondrial complex inhibitors (1.2 µM oligomycin, 1.2 µM carbonyl cyanide p-(trifluoromethoxy)-phenyl-hydrazone (FCCP) and combined 1 µM rotenone with 1 µM antimycin A) were preloaded in the injection ports. For basal rate measurements, ECAR (Extra Cellular Acidification Rate) and OCR (Oxygen Consumption Rate) measurements were assessed. Experiments were performed in triplicates and the data were normalized by cell number.

2.10. Statistical Analysis

All experiments were carried out at least three times unless otherwise indicated. Data were analysed using GraphPad Prism version 7 statistical software (GraphPad Software, La Jolla, CA, USA).

3. Results

3.1. Comparison of PMC42 Cell Lines with Other BC Cell Lines (Luminal, Basal A and Basal B)

Hierarchical clustering of RT-qPCR data of the 9 breast cell lines for EMP markers, inducers and regulators, along with the BC clinically relevant *ESR1*, *PGR* and *ERBB2* gene products, showed substantial variation across the cell lines and revealed two major branches (Figure 1A). Luminal cell lines MCF-7 and T47D, along with the Basal A MDA-MB-468 cell line formed one cluster, whereas the PMC42 cell lines were more closely associated with the other cluster of four Basal B cell lines (BT-549, SUM-159, MDA-MB-157, HS578T). Interestingly, when the heat map was computed on the basis of $\log_2$ fold difference in expression of each gene with respect to PMC42-ET cells, the PMC42-LA cell line clustered more closely with luminal MCF-7, T47D cells and Basal A MDA-MB-468 cells (Figure 1B). This further illustrates that PMC42-LA has more predominant epithelial markers as compared to its parental PMC42-ET cell line. Interestingly, we also observed the complete absence of expression of the FoxA1 gene in PMC42-ET cell line.
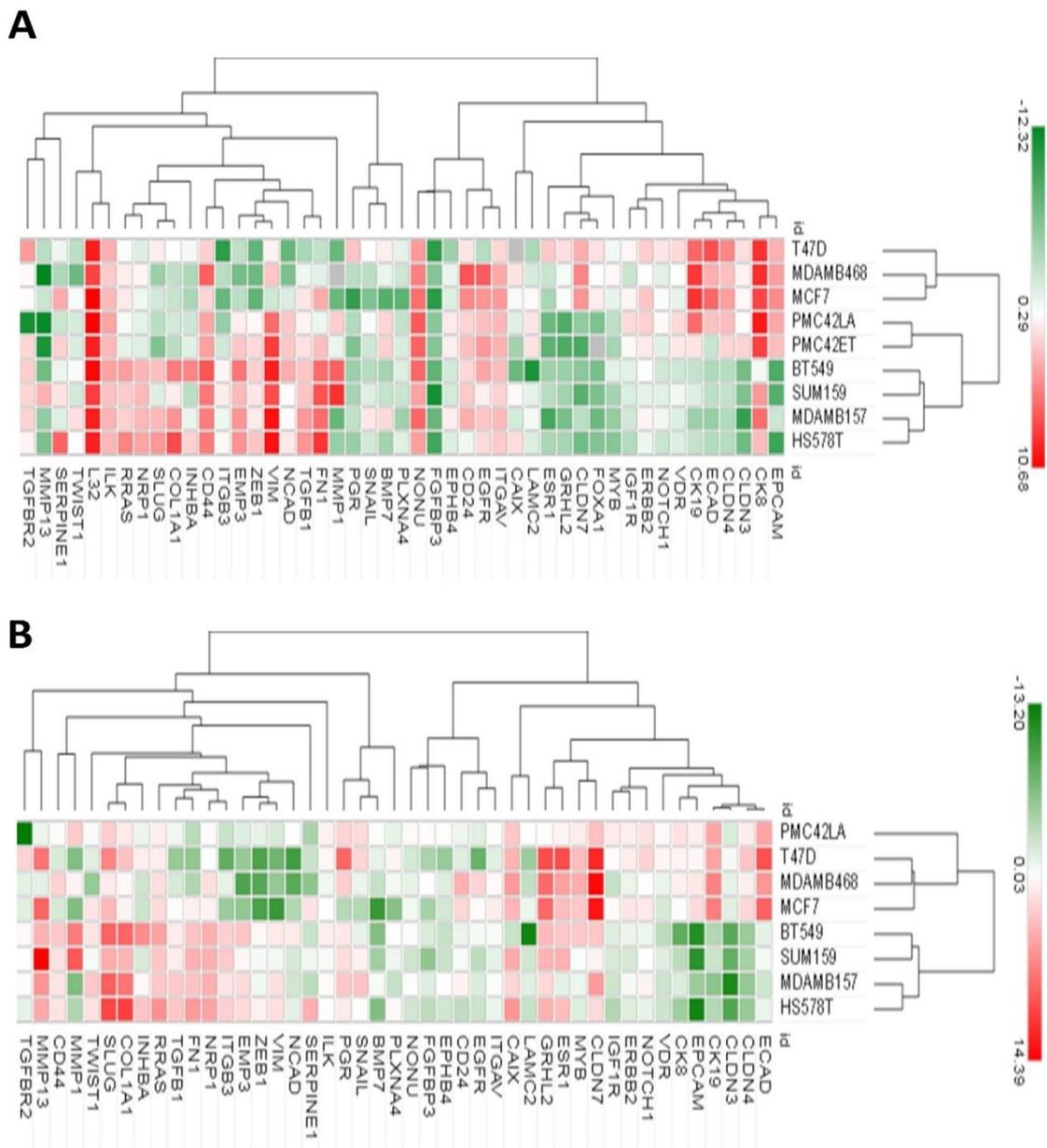
Figure 1. Gene expression heatmaps of selected breast cancer cell lines. (**A**) The normalized mRNA expression values of the breast cancer cell lines obtained against the overall mean expression of all measured genes were subjected to unsupervised hierarchical clustering using Morpheus (Gene-E tool). In the average-linkage cluster algorithm, Pearson correlation was used to measure dissimilarity. (**B**) Unsupervised cluster analysis of the relative expression values of breast cancer cell lines with respect to PMC42-ET.

3.2. CD44$^+$CD24$^{-/low}$ Phenotype Association with Breast Molecular Subtypes and Other EMT Markers

The CD44^{high}/CD24^{low} profile is a putative marker of cancer stemness and is also associated with EMT phenotype [25]. The proportions of CD44$^{+/high}$/CD24$^{-/low}$ cell populations across the PMC42 cell lines were compared with five other breast cancer cell lines (MCF-7, T47D, MDA-MB-468, MDA-MB-231 and SUM-159) and the expression of these surface markers was simultaneously assessed by FACS analysis (Figure 2A,B). Surprisingly, PMC42-LA cells were remarkably stem-like in relation to these markers, as predominantly all cells gated within the CD44^{high}/CD24^{low} subpopulation, similar to MDA-MB-231 and SUM-159 cells representing the most mesenchymal Basal-B subgroup [25], whereas

75.2% of the more mesenchymal PMC42-ET cells were in the CD44^{high}/CD24^{low} state. In agreement with previous reports [64], MCF-7 cultures (Luminal subgroup; [25]), had a small population of CD24$^{-/low}$/CD44^{+} cells (~20%) and the MDA-MB-468 cell line (Basal-A subgroup; [25]), had higher expression of both CD44 and CD24 markers (96.4% of cells gated in CD44^{high}/CD24^{high} state). Compared to MCF-7 and T47D, all other cell lines were mainly constituted by cells with high levels of CD44, and except MDA-MB-468 also showed lower proportions of CD24 expression, consistent with their mesenchymal subgrouping (Figure 2A,B).
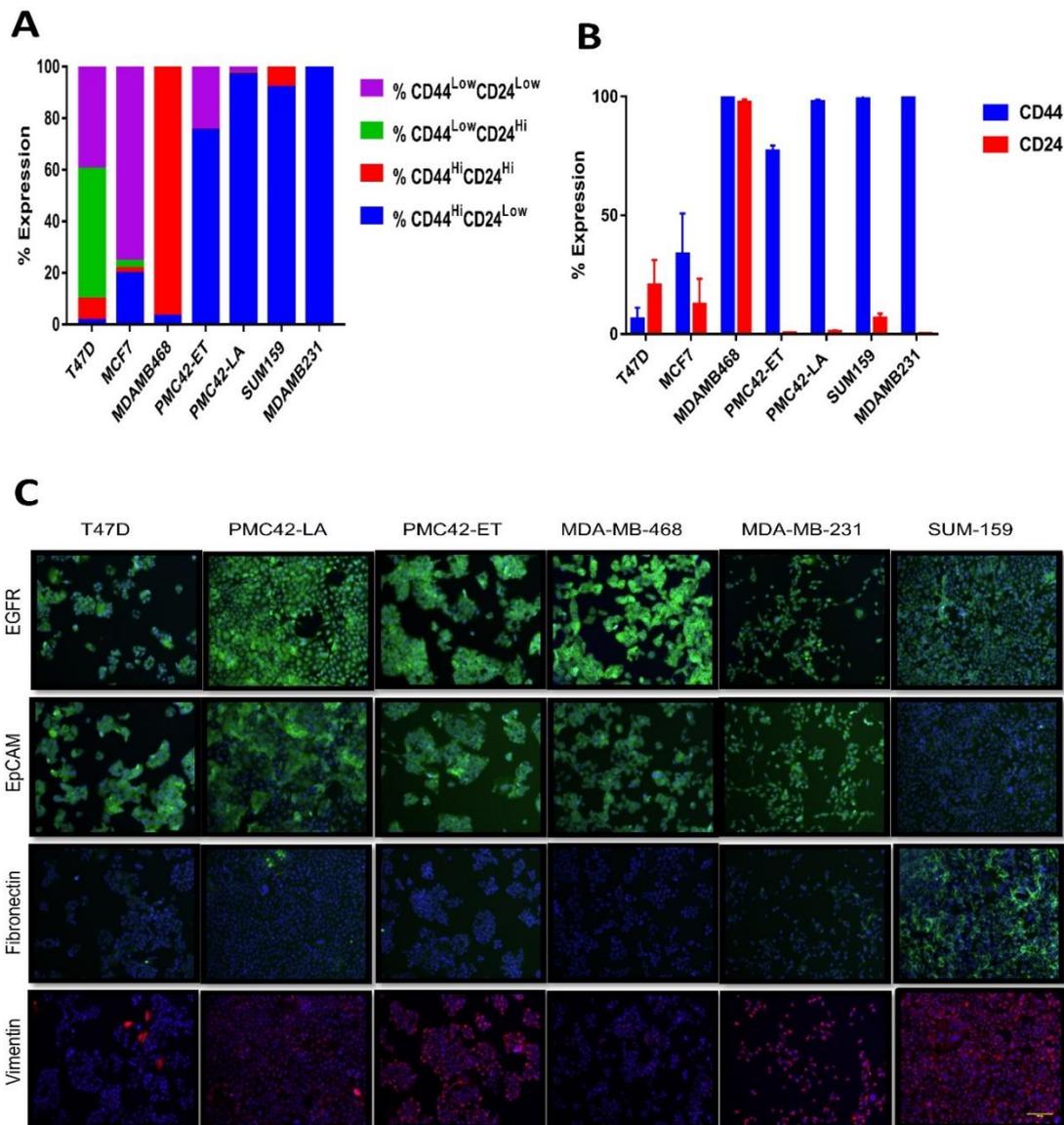

Figure 2. Assessment of stemness and other EMP markers. (**A**) Proportions of the subpopulations defined by the combination of the stem cell markers CD44 and CD24 in PMC42 cell lines and a panel of breast cancer cell lines. (**B**) The relative expression of the stem cell markers CD44 and CD24 in a panel of breast cancer cell lines representative of different molecular subtypes by flow cytometry. Mean ± SD of three independent experiments is shown. (**C**) Immunofluorescence microscopy analysis of EMT marker proteins. Cell lines were stained with antibodies against the epithelial marker EpCAM and against the mesenchymal markers EGFR, vimentin and fibronectin. Scale bar, 100 μm.

The CD44$^{+/high}$/CD24$^{-/low}$ cell populations across PMC42-ET, PMC42-LA, MCF-7 and MDA-MB-468 were also assessed after stimulation of the cells with EGF or TGF-β for 72 h. Both MCF-7

and MDA-MB-468 cell lines exhibited a 10%–20% increase in their CD44^{high}/CD24^{low} proportions after treatment with EGF, but not with TGF-β. PMC42-ET cells, however, could be made potentially stem-enriched with 93%–94% population of cells in CD44^{high}/CD24^{low} state after stimulation with either EGF or TGF-β, which was due to increased CD44 expression (Supplementary Figure S1).

EMT-associated markers EGFR, EpCAM, fibronectin and vimentin were also evaluated using immunofluorescence for the six breast cancer cell lines (Figure 2C). EGFR expression was highest for MDA-MB-468 cells, in which EGFR is known to be amplified [65]. EpCAM expression at cellular junctions was observed for T47D, PMC42-LA and MDA-MB-468 cell lines, and fibronectin expression was only observed for SUM-159 cells. Vimentin expression was universally positive in SUM-159, MDA-MB-231 and PMC42-ET cells, whereas on an average 10% and 20% of the cells were positive for vimentin expression in MDA-MB-468 and PMC42-LA cells, respectively, as quantified using IN Cell Investigator software, which is consistent with previous reports [39,40,66].

3.3. Comparative RNA-seq Analysis of PMC42 Cell Lines

RNA-seq results obtained previously for the EGF-induced EMT studies in PMC42 system [30] were re-analysed to study the transcriptional differences between the two PMC42 cell lines. Comparative analysis was investigated using the "Hallmark" geneset collection within the MSigDB of GSEA and IPA. Negative enrichment for signatures related to EMT ($p < 0.001$, NES = −1.73), TNFA signalling via NFκb ($p = 0.007$, NES = −1.55), inflammatory response signature ($p = 0.004$, NES = −1.52) and hypoxia signature ($p = 0.016$, NES = −1.41) were observed in PMC42-LA with respect to its parental cell line PMC42-ET using GSEA (Figure 3A). Using IPA for the comparative RNA-seq analysis, the top-five significant upstream regulators we identified were focused on the inhibition of TNF, TGFβ-1, EGFR and JNK gene in PMC42-LA, whereas estrogen receptor gene was considered an activated regulator in PMC42-LA. IPA also reported a significant gene network indicating the importance of TWIST2 downregulation and SPDEF upregulation in the epithelial PMC42-LA cells (Supplementary Figure S2A).

3.4. Comparative Proteome Quantification of Alterations in the PMC42 Cell Line System

We next performed comparative proteomics and subjected protein extracts from the two PMC42 cell lines to mass spectrometry. Among a total of 2460 identified and annotated proteins in the PMC42 cell lines, 244 proteins were expressed at significantly different levels in the two cell lines (adjusted p value < 0.01). Of these, 73 proteins were significantly upregulated, and 61 proteins were significantly downregulated by a factor of 2-fold or more in the epithelial PMC42-LA cells. KEGG pathway analysis indicated that differentially regulated proteins were involved in glycolysis/gluconeogenesis (*ALDH1A3, ALDH3A1, ALDOC*) ($p = 0.00073$), proteasome (*PSB3, PSB6, PSB7, PSMD1-4*) ($p = 6.06 \times 10^{-8}$), protein processing in endoplasmic reticulum ($p = 3.06 \times 10^{-6}$) and carbon metabolism (*ALDOC, SUCA, G6PD, HXK1, DLDH*) ($p = 0.00039$), respectively. The volcano plot that shows the difference in protein levels between PMC42-ET and PMC42-LA also highlights several EMT markers, among which mesenchymal markers, such as VIM and EGFR, were upregulated by 4-fold ($p < 2.51 \times 10^{-5}$) and 2-fold ($p < 5.83 \times 10^{-5}$), respectively, in PMC42-ET cells (Figure 3B). Epithelial markers, such as KRT19 and F11R (Junctional adhesion molecule A), were significantly upregulated in PMC42-LA by 4.75-fold ($p < 0.00066$) and 4.6-fold ($p < 00093$), respectively. IPA also deduced glycolysis (with a z-score of 3.5), aryl hydrocarbon receptor signalling (with a z-score of 3.1) and ILK signalling (with a z-score of 1.6) as significantly upregulated canonical pathways in the epithelial PMC42-LA cell line. Gene network enrichment plot from proteomics analysis indicated a possible role of NFκb complex dysregulation in PMC42-LA cells (Supplementary Figure S2B), which is also in agreement with the similar GSEA findings from RNA-seq analysis.
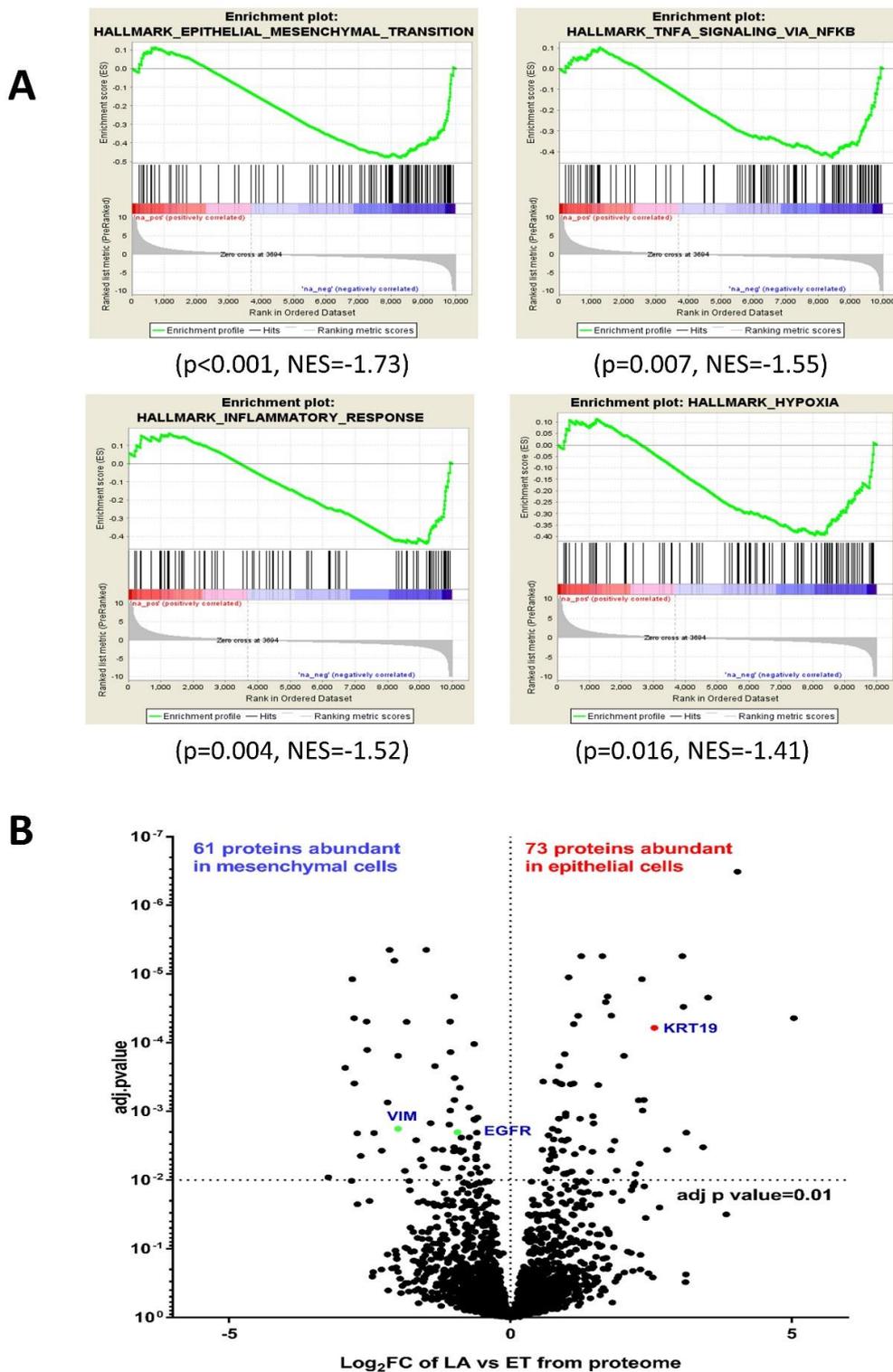
(p<0.001, NES=-1.73)   (p=0.007, NES=-1.55)

(p=0.004, NES=-1.52)   (p=0.016, NES=-1.41)

Figure 3. Comparative analysis of transcriptomic and proteomic data from PMC42 cell lines. (**A**) Results from representative enrichment plots from Gene Set Enrichment Analysis (GSEA) ($p < 0.05$) are shown from comparative transcriptome analysis. These data revealed a significant negative enrichment for gene sets involved in EMT in PMC42-LA. (**B**) Quantitative proteome analysis of PMC42-LA in comparison to PMC42-ET reflected in volcano plot shows that 73 proteins were significantly upregulated, and 61 proteins were significantly downregulated.

3.5. Karyotypic Heterogeneity Exists within and across the Sister Breast Cancer Cell Lines PMC42-ET and PMC42-LA

Chromosomal instability (CIN), including numerical CIN (resulting in aneuploidy) and structural CIN (resulting in partial chromosomal gains and losses and translocations) are inherent in cancer and underpin many of the phenotypic manifestations that contribute to cancer progression. Hence, in order to assess their possible contributions to the spontaneous MET seen in the PMC42 system, we explored whether karyotype differences existed in the PMC42 cell lines using metaphase spreads and karyotypic G-banding.

First, we determined the copy number status per chromosome. Chromosomal counts from individual cells in each PMC42 cell line were plotted as a heatmap, where PMC42-ET and PMC42-LA cells clearly are seen as two separate clusters (Figure 4A). CIN, and resulting aneuploidy, is a hallmark of cancer, and despite this strong partition into 2 clusters, variable ploidy distribution was also reflected within each of the PMC42 cell lines. We identified copy number differences of individual chromosomes (Figure 4A). Significant differences in the ploidy levels between the cell lines are represented in Table 1, where eight chromosomes show high chromosomal number differences across the two cell lines. Tetrasomy was observed more often for chromosomes 3, 5, 7, 19 and 22 in PMC42-ET, whereas in PMC42-LA, tetrasomy was observed for chromosomes 7 and 19. Loss of chromosome 22 was found in 100% of PMC42-LA cells, as only a single copy was present in each of 50 cells assessed for karyotyping. The patterns of positive and negative chromosomal correlation with regards to their ploidy were also studied for individual cells in the two PMC42 cell lines, where no strongly significant association was confirmed for PMC42-LA, but positive correlations were observed between chromosomes 9 and 14, and 17 and 20, respectively ($r^2 = 0.67$ and 0.59, Pearson correlation) (p-value of 7.64×10^{-8} and 6.37×10^{-6}), and a negative correlation ($r^2 = -0.51$) (p-value of 0.0001) was found between chromosomes 9 and 12, in the PMC42-ET cell line (Supplementary Tables S3 and S4). Overall, PMC42-ET was primarily comprised of near-triploid karyotypes with a modal number of 68 chromosomes (range 59–75), whereas PMC42-LA was primarily comprised of near-triploid karyotypes with a modal chromosome number of 63 (range 52–64) (Figure 4B). The total chromosome numbers in the two cell lines are significantly different ($p < 0.0001$) (Figure 4C). Two observations suggest an increased level of CIN in PMC42-ET compared to PMC42-LA. First, PMC42-ET cells show a broader range of chromosome numbers, 17, compared to 13 in PMC42-LA cells and more individual chromosome numbers deviate from the modal number (Supplementary Figure S3A,B). Second, GSEA on our RNA-seq data shows a significant depletion in the expression of 70 genes that are part of the well-established CIN70 signature [67] in PMC42-LA cells compared to PMC42-ET cells ($p = 0.027$; Figure 4D). This suggests that CIN may have promoted the transition from PMC42-ET to PMC42-LA.

Next, as expected, karyotyping revealed that the PMC42-LA derivative cell line harbours some of the major structural rearrangements seen in the parental PMC42-ET cells. For example, one of the arms of each of chromosomes 2, 3 and 8 is shorter or truncated, the *p*-arm of chromosome 9 is fused with the long arm of chromosome 10, and chromosome 21 has a third copy of its *p*-arm fused with the long arm of chromosome 7 (Figure 5A,B). Some characteristics, such as truncated arms of chromosomes 2 and 3, trisomy 1 and trisomy 20, and a modal chromosomal number of 66, were consistent with initial reports in 1983 [42]. In PMC42-LA cells only, we also observed a few (in the range of 1–4) marker chromosomes whose derivative chromosomal origins cannot be recognized via karyotyping (Figure 5B, marked as 'mar' and Figure 5C, marked as 'UNC'). Ploidy distributions of each chromosome from 50 karyotyped cells of each cell line also reflect the dynamics of copy number alterations at the chromosomal level (Figure 5C).
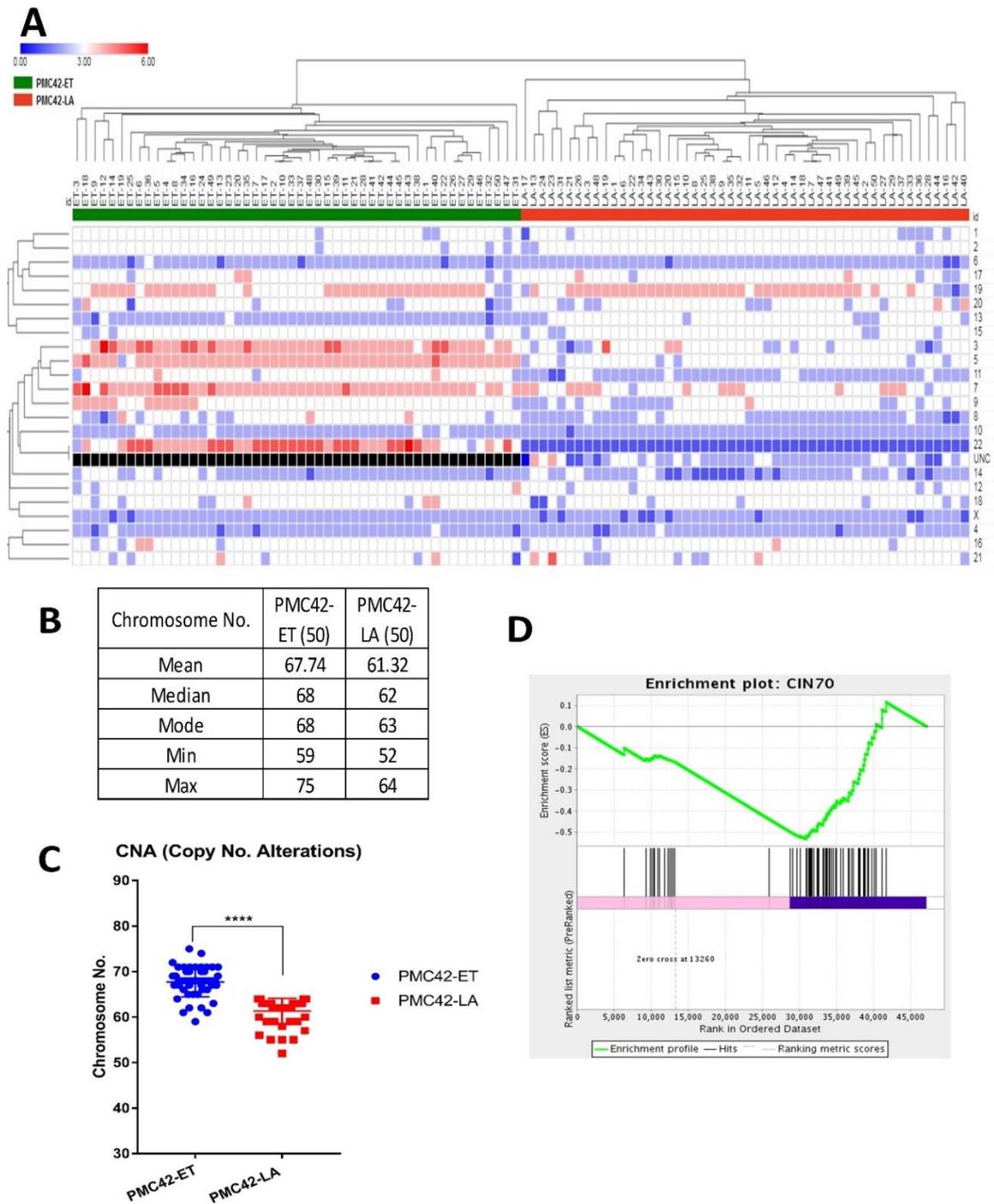
Figure 4. Chromosomal ploidy distribution of PMC42 cell lines. (**A**) Heatmap for copy number distribution of chromosomes deciphered from 50 karyotypes from each of the PMC42-ET and PMC42-LA cell lines *(UNC: Unidentified Chromosome)*. (**B**) Distribution of chromosome numbers of PMC42-ET and PMC42-LA cell lines. (**C**) Chromosome numbers analysed from PMC42-ET and PMC42-LA for a total of 50 cells were compared for each cell line. Significance was determined by an unpaired *t* test with Welch's correction, with **** $p < 0.0001$. (**D**) CIN70 enrichment plot following Gene Set Enrichment Analysis.

Table 1. Ploidy alterations of 50 single cells from PMC42-ET and PMC42-LA were compared using *t*-test.

Chromosome No.	*p*-Value across ET vs. LA
22	5.28847×10^{-45}
5	1.50398×10^{-28}
13	7.71473×10^{-23}
11	9.42164×10^{-20}
3	3.33503×10^{-18}
7	8.49205×10^{-14}
8	3.73987×10^{-8}
9	1.08547×10^{-6}
10	0.012162341
X	0.0151172
14	0.022502942
18	0.083804992
12	0.260603283
1	0.278286015
15	0.531258862
2	0.678929758
6	0.717332498
16	0.748135128
4	0.75812698
17	0.771687988
20	0.814301536
21	0.823855275
19	0.928168854

Taken together, these results indicate that numerical chromosomal heterogeneity exists between and within the PMC42 cell lines. In addition, PMC42-ET cells show features of CIN and PMC42-LA cells harbour some specific structural abnormalities not observed in PMC42-ET cells. These changes could underpin the transition of PMC42-ET cells to a more epithelial phenotype.

3.6. Cancer Driver Mutations in PMC42 Cell Lines

WES was performed to probe more deeply the genetic aberrations in the PMC42 system. After applying selective filters for delineating deleterious mutations within exons, we identified 465 SNVs in PMC42-ET and 475 SNVs in PMC42-LA (Figure 6A,B). We considered missense, non-sense, frame-shift and splice site mutations that involve structural and functional alteration of the protein products as deleterious. The number of Indels present in PMC42-ET were 83 and in PMC42-LA were 85. Approximately 75.4% of the somatic mutations and 60% of the Indels were shared between the parental and derivative cell line (Figure 6C,D). The results of deleterious SNV and Indels identified using WES across PMC42-ET and PMC42-LA cells are shown in the supplementary document (Supplementary Tables S5 and S6 for PMC42-ET and Tables S7 and S8 for PMC42-LA). CHASM score was computed for all the missense mutations identified in the PMC42 cell lines to identify driver mutations. The top 9 potential drivers that were common between the two PMC42 cell lines were *TP53, MERTK, DNMT3A, CPZ, PPM1H, PPIP5K2, C10orf76, DNAH7, CFTR* (Figure 6E). We also compared the driver mutations identified in the PMC42 cell lines with the TCGA mutations dataset using the CRAVAT interface (Figure 6F). The TP53 mutation site (H36R) was reported earlier in the TCGA dataset, whereas the other top 4 driver missense mutation locations identified were considered novel, as they were not reported in the TCGA dataset (as observed from the CRAVAT interface). The deleterious genes

identified in the PMC42 cell lines were also stratified in a gene-family matrix according to their known role in cancers using the Broad Institute's GSEA analysis. We identified three commonly mutated tumour suppressor genes between two PMC42 cell lines (*NF2 (attaining stop gain function), TP53, TSC2)* and an additional mutated tumour suppressor gene *ATM* in PMC42-ET only. Eight commonly mutated oncogenes (*ARNT (TF), EML4, GNAS, NTRK3 (PK), PER1 (TF), TCL1A (TF), TLX1, TTL)* and 2 additional mutated oncogenes *TAL1 and IL6ST* in PMC42-ET only were also identified for their role in tumorigenesis (Table 2; Table 3). Interestingly, the gene-family matrix derived for the deleterious genes in the PMC42-LA cell line in reference to parental PMC42-ET cell line identifies mutations in 2 significant EMT-promoting transcription factors, *SNAI2* (K188N) and *SOX3* (R22P) (Supplementary Tables S9 and S10) and other EMT genes, such as *GSN (R397W), WNT1 (T363K), ITGA4 (R565W) and NID2 (G426E)* (Supplementary Table S7). In addition to that, the EMT-associated splice variation regulator *ESRP2* (R248S) was found to be mutated in the parental PMC42-ET cell line.
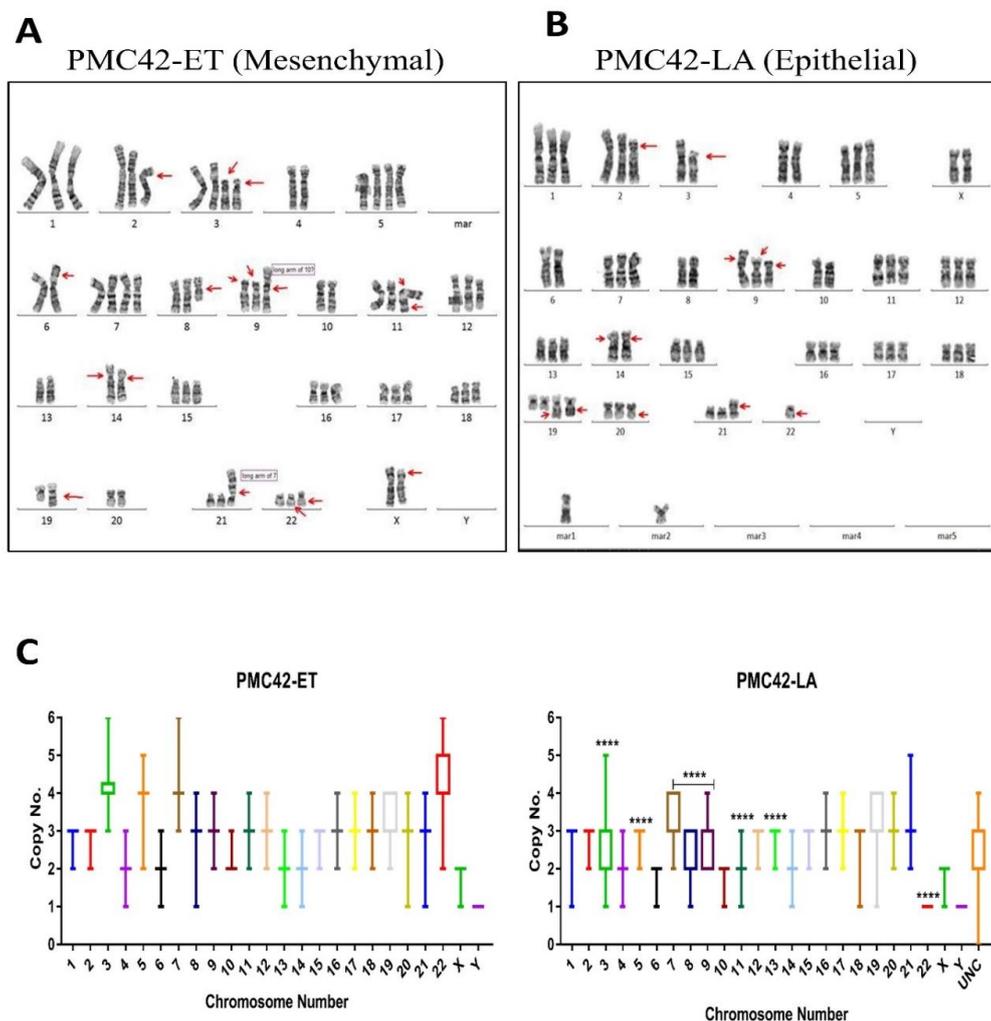

Figure 5. Karyotypic analysis of PMC42 cell lines. (**A**) A representative G-banded karyotype of the near-triploid cell line PMC42-ET, showing structural and numerical changes. (**B**) A representative G-banded karyotype of the cell line PMC42-LA. *Arrows* point to main chromosomal alterations. *Mar$_n$* marker chromosome (**C**) Ploidy distribution of each chromosome is presented for PMC42-ET and PMC42-LA from 50 karyotyped cells. *P*-values are indicated (as described in Table 1 using Student's *t*-test), and data presented in box (median, first and third quartiles) and whisker (extreme value) plots (*UNC: Unidentified Chromosome*).
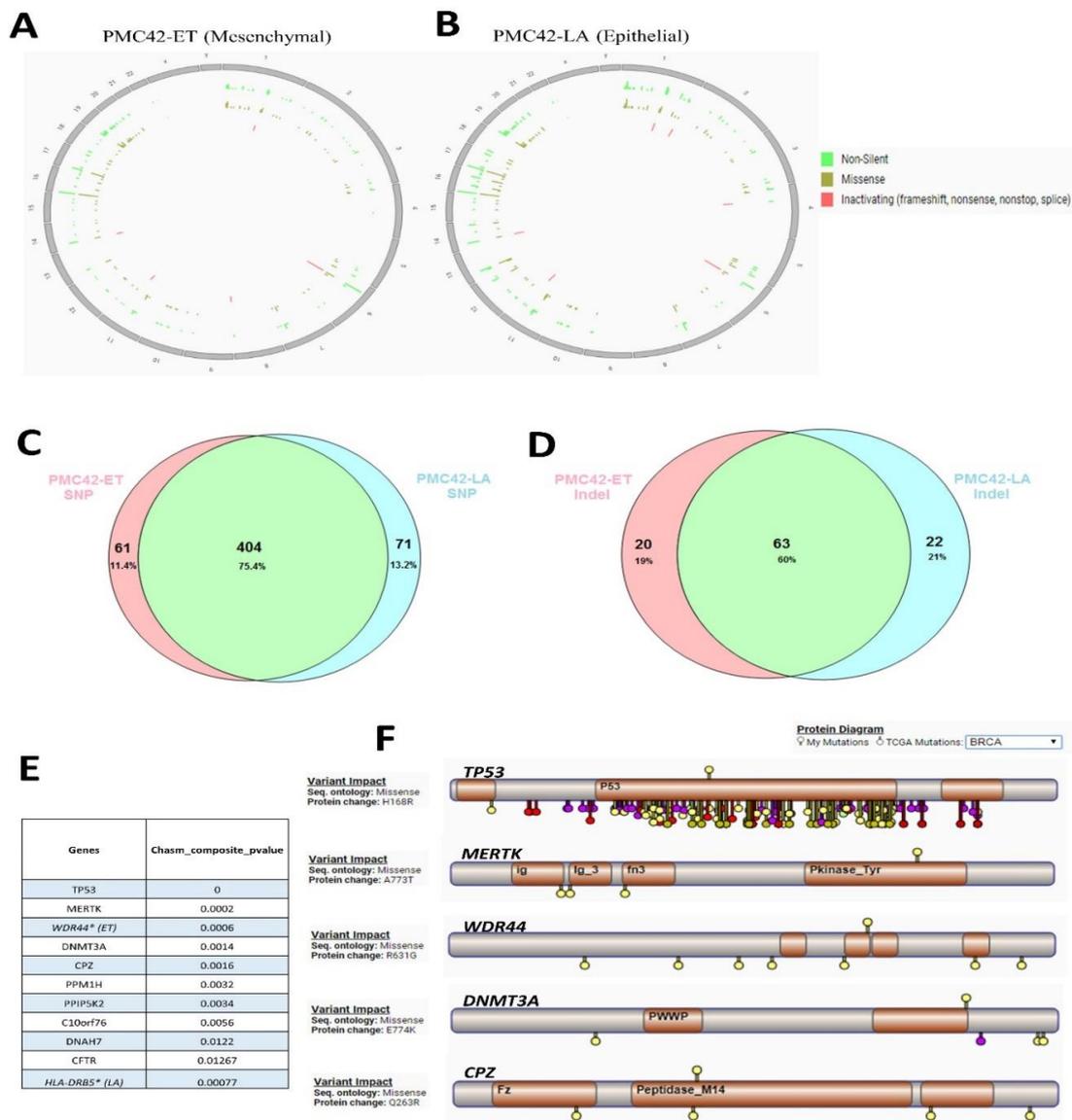
Figure 6. Assessment of whole-exome sequencing (WES) from PMC42 cell lines. Chromosome-specific distribution of non-silent, missense, and inactivating mutations are displayed on Circos plots for PMC42-ET (**A**) and the derivative PMC42-LA (**B**) cell lines, respectively. Representation of shared and unique (**C**) SNVs and (**D**) indels discovered by WES for PMC42-ET and PMC42-LA. (**E**) CHASM score was computed and (**F**) the top 10 potential driver mutations for PMC42-ET and PMC42-LA were determined. The top 5 potential driver mutations for PMC42-ET and PMC42-LA were annotated for their presence within protein sequences (indicated on the top of each gene) and compared for somatic mutations identified in TCGA dataset (indicated on the bottom of each gene) for the same protein sequences, using the software CRAVAT [68].

Table 2. Gene set enrichment analysis (GSEA) for genes with deleterious mutations (SNV and Indels) identified in PMC42-ET.

	Cytokines and Growth Factors (CGF)	Transcription Factor (TF)	Homeodomain Proteins (HP)	Cell Differentiation Markers (CM)	Protein Kinases (PK)	Translocated Cancer Genes (TCG)	Oncogenes	Tumour Suppressors
Tumour suppressors	0	1	0	0	1	0	0	4*
Oncogenes	1	4	1	1	1	8	10&	
Translocated cancer genes	0	4	1	0	1	8*		
Protein kinases	0	0	0	0	17#			
Cell differentiation markers	2	0	0	8!				
Homeodomain proteins	0	4	4@					
Transcription factor	0	28$						
Cytokines and Growth Factors	6^							

Note: (**4* tumour suppressor genes** are ATM (PK), NF2, TP53 (TF), TSC2 **8* translocated cancer genes/oncogenes** are ARNT (TF), EML4, NTRK3 (PK), PER1 (TF), TAL1 (TF), TCL1A (TF), TLX1, TTL and **2& additional oncogenes** are GNAS, IL6ST (CM and CGF) **17# protein kinases** are ANKK1, ATM, AURKB, CSNK1G2, FRK, LRRK2, MAP3K9, MAST4, MERTK, NRBP2, NTRK3, PASK, PKMYT1, PRKDC, ROCK2, SLK, TTN **8! cell differentiation markers** are CR1, CR2, FCGR3A, IGLL1, IL6ST, LILRB5, MSR1, SEMA7A **4@ homeodomain proteins** are HOMEZ, IRX3, TLX1, ZEB2 **28$ transcription factors** are ARNT, CEBPZ, CHD4, E2F1, ESR1, EYA3, FOXI1, HOMEZ, IRX3, MED21, NEUROD4, NR1I2, PER1, PRDM2, RFX5, RREB1, RRN3, SP1, SRA1, SUPT5H, TAL1, TFDP3, TLX1, TP53, UHRF1, ZEB2, ZNF160, ZNF91 **6^ cytokines and growth factors** are C5, CMTM2, CYR61, IL6ST, SEMA6D, SEMA7A) (Green colour denotes mutations unique to this cell line).

Table 3. Gene set enrichment analysis (GSEA) for genes with deleterious mutation (SNV and Indels) identified in PMC42-LA.

	Cytokines and Growth Factors	Transcription Factor	Homeodomain Proteins	Cell Differentiation Markers	Protein Kinases	Translocated Cancer Genes	Oncogenes	Tumour Suppressors
Tumour suppressors	0	1	0	0	0	0	0	3*
Oncogenes	0	3	1	0	1	7	8&	
Translocated cancer genes	0	3	1	0	1	7*		
Protein kinases	0	0	0	0	15#			
Cell differentiation markers	1	0	0	6!				
Homeodomain proteins	0	3	3@					
Transcription factor	0	24$						
Cytokines and Growth Factors	5^							

Note: (**3* tumour suppressor genes** are NF2, TP53 (TF), TSC2 **7* translocated cancer genes/oncogenes** are ARNT (TF), EML4, NTRK3 (PK), PER1 (TF), TCL1A (TF), TLX1, TTL and 1& additional oncogene is GNAS, **15# protein kinases** are ANKK1, AURKB, CSNK1G2, FRK, MAP3K9, MAST4, MERTK, NRBP2, NTRK3, PASK, PKMYT1, RIPK2, ROCK2, RPS6KA5, STK16 **6! cell differentiation markers** are CR1, CR2, FCGR3A, IGLL1, LILRB5, SEMA7A **3@ homeodomain proteins** are HOMEZ, TLX1, ZEB2 **24$ transcription factors** are ARNT, CEBPZ, CHD4, E2F1, ESR1, FOXI1, HOMEZ, MED21, NEUROD4, NR1I2, PER1, PRDM2, RFX5, RREB1, RRN3, SP1, SRA1, SUPT5H, TLX1, TP53, UHRF1, ZEB2, ZNF160, ZNF318 **5^ cytokines and growth factors** are C5, CMTM2, CYR61, SEMA6D, SEMA7A (Green color denotes mutations unique to this cell line).

3.7. Inference of CNV from Exome Sequencing Data

The estimated copy number ratios of chromosomal segments for PMC42-ET relative to PMC42-LA are shown in Figure 7. Control-FREEC was used to determine the copy number ratio profiles and to identify regions with significant amplification or loss. The copy number profiles deduced from WES resulted in a total of 166 gain and 34 loss segments in PMC42-ET, relative to PMC42-LA. The most significant losses are from 5p, 20p, Xq and whole chromosome 13, while the major gains are in chromosomes 3, 5q, 7q, 10q, 11q, 20q and 21. The identified genomic regions of amplification or loss were also consistent with our karyotyping studies. Copy number changes of the regions (i.e., amplification of chromosomes 3, 5, 7, 9,11, 22 and loss of chromosomes 13 in PMC42-ET relative to PMC42-LA) detected in the WES study were also identified in karyotyping. Additionally, WES data reflects amplification in chromosomes 10q and 21, which might be due to nonreciprocal translocation events of an additional chromosome 9 with chromosome 10q and chromosome 21 with 7q as shown in karyotype analysis. WES helps in revealing the amplifications/losses at the gene level, whereas karyotyping reflects the overall ploidy distribution better (as reflected in Figure 5), as it is a representation of single cells. The amplified/lost regions from WES data for PMC42-ET with respect to PMC42-LA is also provided in additional Supplementary Table S11.

3.8. TGFBR2 Ablation and Influence on EMT Induction in PMC42-LA

When analyzing the segments of genes that were completely lost in either of the PMC42 cell lines, we identified two regions from PMC42-LA that were completely missing. The chromosome 1 region carrying genes *FOXD2, FOXE3* and the chromosome 3 region containing genes *GADL1, RBMS3, TGFBR2* were completely lost from the PMC42-LA cell line. The specific functions of *FOXD2* and *FOXE3* genes are yet to be determined, however the role of *TGFBR2* in EMT induction is well established [27,69,70]. The surface expression of TGFβR2 in the PMC42 system was analysed in comparison with several other cell lines (T47D, MDA-MB-468, SUM-159 and MDA-MB-231) using FACS (Figure 8A). High TGFβR2 surface expression was seen in almost all the cells for Basal B cell lines SUM-159 and MBA-MB-231, whereas only 40% of PMC42-ET cells expressed TGFβR2 on their surface, and TGFβR2 expression was completely absent on PMC42-LA cells, which is in concordance with our WES-deduced results. The surface expression of TGFBR2 on T47D cells was also negligible, consistent with their low EMT-associated TGF-β enrichment score (TES) [27]. RNA-seq analysis of the PMC42 cell lines was also interrogated using the algorithms as described [27] to obtain their TES values and identify any evidence of intrinsic TGF-β-induced EMT. The PMC42-ET had a TES value of 0.594 which is relatively high compared to the PMC42-LA TES value of −0.015. The low/negative TES value of PMC42-LA is also in concordance with the previously deduced TES values from Luminal and Basal A cell lines (MCF7: −0.58977; T47D: −0.69277; MDAMB468: −0.66892), whereas Basal B cell lines have relatively higher TES values (MDA-MB-231 0.130999; SUM159PT: 0.430126). TES values for the various breast cancer cell lines are taken from Supplementary data file S10 of [27].

PMC42-ET and PMC42-LA cell lines were also tested for their EMT induction with EGF, TGF-β and combined treatments using RT-qPCR (Figure 8B). E-cadherin was significantly downregulated with combined EGF and TGF-β growth factor treatments in the PMC42-ET cell line compared to either EGF or TGFβ alone, suggesting that TGF-β augments the previously reported EGF-induced EMT [30]. There was significant upregulation of mesenchymal markers vimentin, Slug and CD44 in both the cell lines with EGF, and with combined EGF and TGF-β treatment in PMC42-ET cells, however there were no effects of TGF-β treatments on PMC42-LA cells (Figure 8B,C). At an individual factor level, assessed via transcriptomics, several modulators or mediators of TGF-β signalling were impacted across PMC42 cell lines (e.g., AGR2, RhoA, TGFB1, CTNNB1, JUNB) were significantly downregulated in PMC42-LA. Thus, the PMC42-LA cell line did not display any predisposition to undergo EMT-like changes in mesenchymal gene expression with TGF-β treatment.
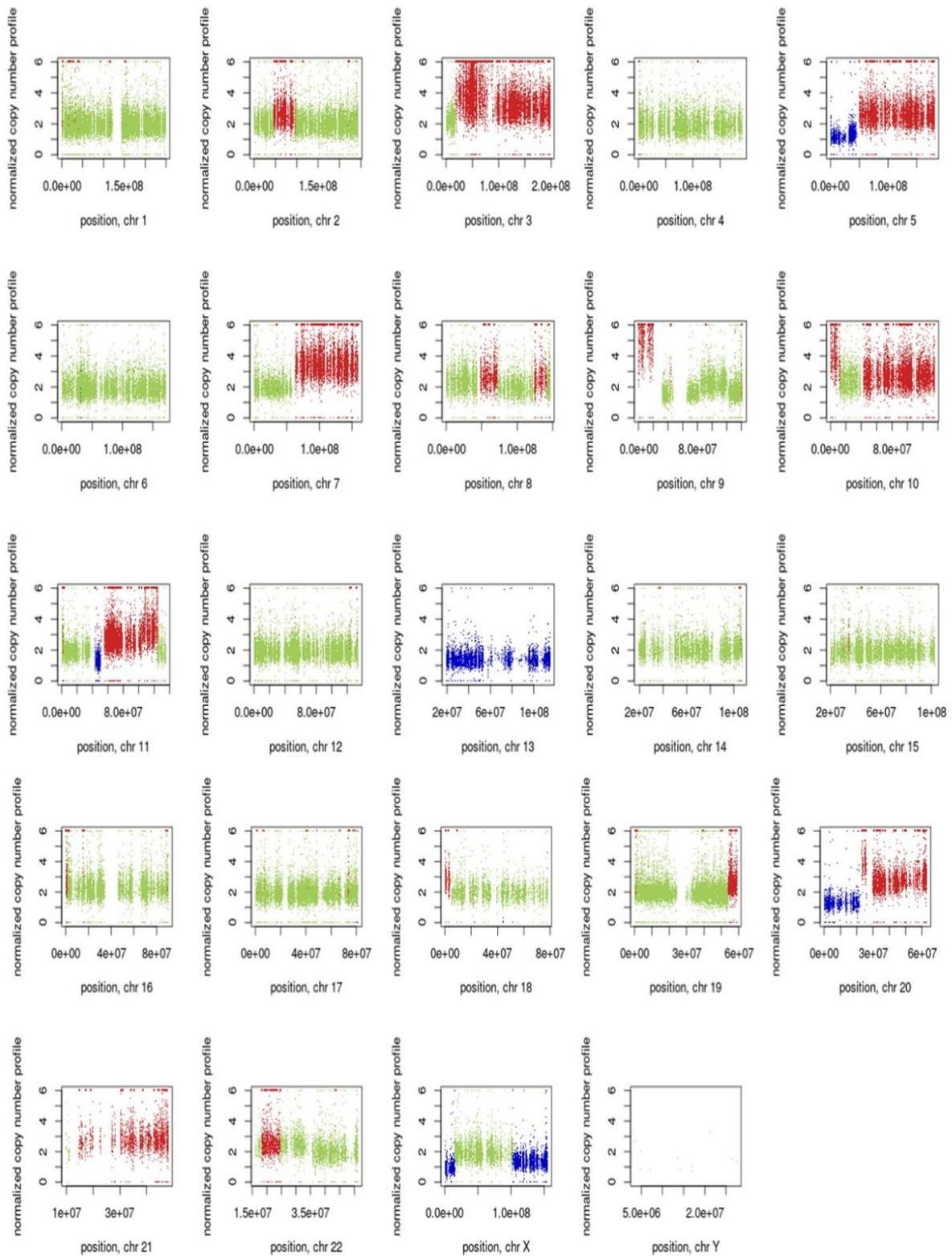
Figure 7. Visualization of Control-FREEC v 6.0 output from PMC42 cell lines exome sequencing data (Illumina HiSeq 2000). Copy number profiles for all chromosomes are shown of PMC42-ET in comparison to PMC42-LA, normal copy number status is shown in green, copy number gains are reflected in red and copy number losses are shown in blue.
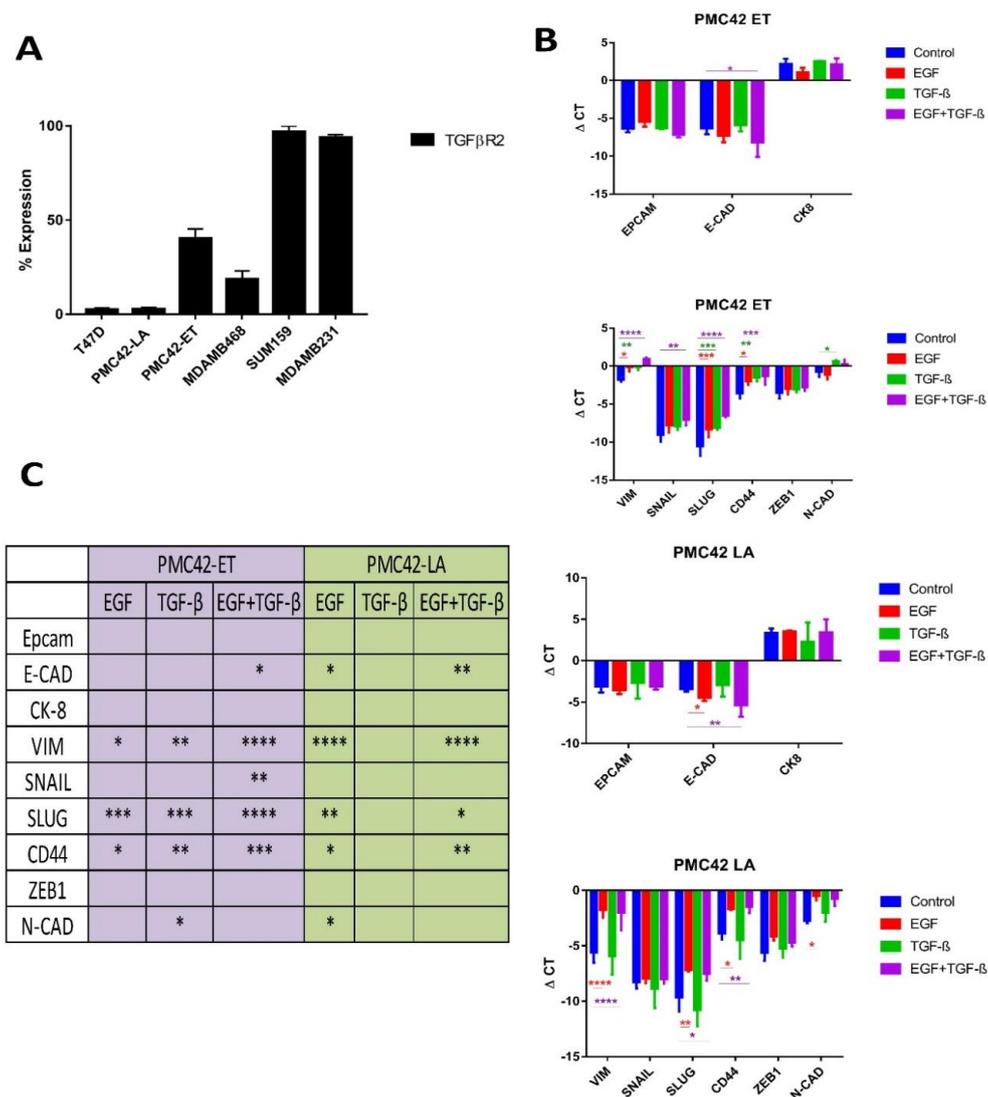
Figure 8. TGFBR2 ablation and influence on EMT induction in PMC42-LA. (**A**) Cell surface expression levels of TGFβR2 in a panel of breast cancer cell lines representative of distinct molecular subtypes. (**B**) PMC42-ET and PMC42-LA cells were treated for 5 days with 10 ng/mL EGF, 10 ng/mL TGF-β and combined 10 ng/mL of EGF and TGF-β. qPCR analysis of epithelial and mesenchymal markers were tested after EMT induction for 6 days with growth factor treatments. dCt values normalized against L32 and as an average from triplicates are shown. Statistical method applied is a two-way ANOVA with * indicating a *p*-value < 0.1, ** *p*-value < 0.01, *** *p*-value < 0.001 and **** *p*-value < 0.0001. (**C**) *p*-values calculated using 2-way ANOVA against each gene expression are tabulated.

	PMC42-ET			PMC42-LA		
	EGF	TGF-β	EGF+TGF-β	EGF	TGF-β	EGF+TGF-β
Epcam						
E-CAD			*	*		**
CK-8						
VIM	*	**	****	****		****
SNAIL			**			
SLUG	***	***	****	**		*
CD44	*	**	***	*		**
ZEB1						
N-CAD		*		*		

3.9. Inter-Data Relationships from CNV and RNA-seq with Proteome Data

Correlations were assessed between protein expression and gene expression, and between protein expression and copy number variation, after applying the filter to only those proteins (*n* = 244) that were significantly dysregulated. For the RNAseq-to-proteome comparison, the Spearman correlation coefficient was 0.748 (Figure 9A), whereas the correlation was only 0.39 for CNV deciphered from WES at the gene level compared to the proteome (Figure 9B). Assessment of impact of CNV changes at whole differential RNA expression level reflects a Spearman's correlation coefficient of $r = 0.002805$, with a non-significant p value of 0.77 (Supplementary Figure S4). Undoubtedly, some of the gene expression levels are influenced by changes in amplification or depletion of gene dosage at allele level, but this is probably masked by the likelihood that gene expression can be significantly modulated by

other factors, such as epigenetics or transcription factors. Interestingly, when extrapolating RNA-seq and proteome results for the 244 differential expressed proteins, AGR2 shows relatively very high gene expression in PMC42-LA compared to PMC42-ET. An essential role of induced AGR2 in re-acquisition of epithelial markers has been reported [71], where activated Smad and Erk signalling cascades were identified as mutually complementary pathways responsible for TGF-β-mediated inhibition of AGR2. Since TGFBR2 expression is absent in PMC42-LA, AGR2 may be playing a crucial role in maintaining the epithelial phenotype of PMC42-LA cells. This is consistent with a strong enrichment of ARG2 expression in the Luminal subgroup of breast cancer cell lines, and some enrichment in Basal A, compared to Basal B (Supplementary Figure S3C).
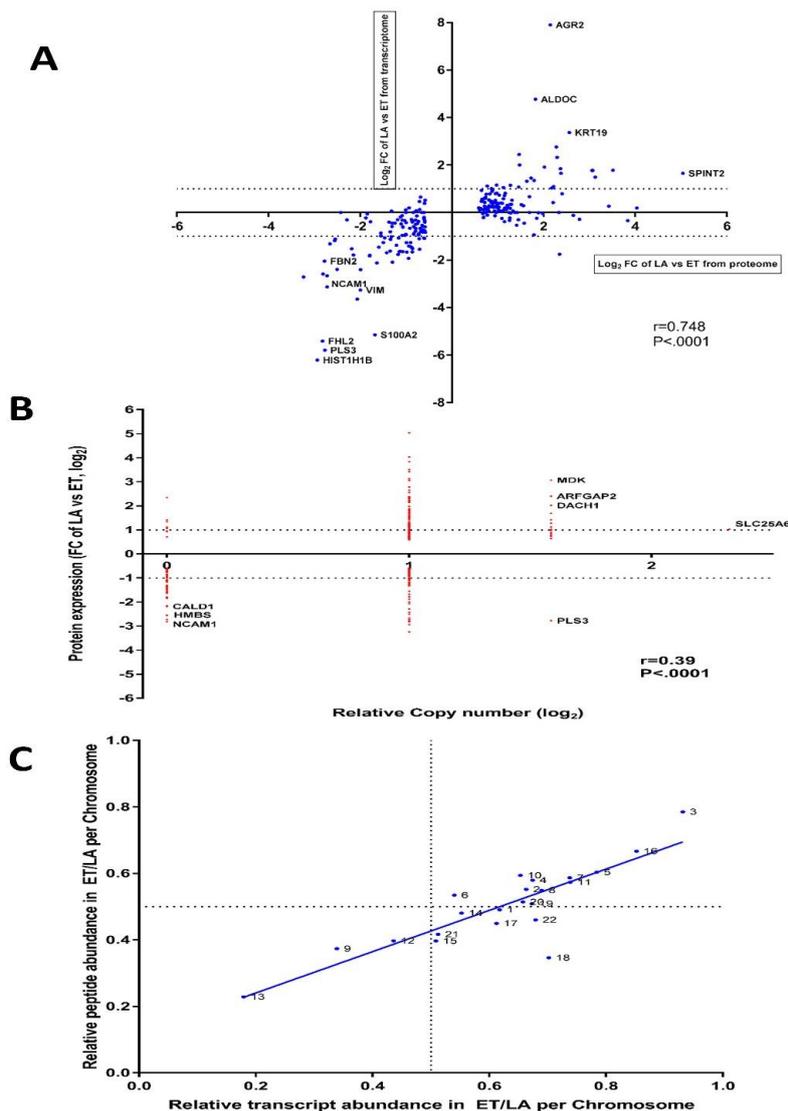
Figure 9. Inter-data relationships from CNV, RNA-seq and proteome studies in PMC42 cell lines. (**A**) Log$_2$ fold change of mRNA and protein expression levels of 244 significantly differential expressed proteins for PMC42-LA vs PMC42 ET were computed. Spearman's correlation coefficient (*r* with *p*-value) between log$_2$ fold change of protein and mRNA expression is indicated at the bottom right. Dotted horizontal bars indicate 2-fold upregulation and downregulation on the log$_2$ scale for mRNA expression. (**B**) Log$_2$ fold change of 244 significantly differential expressed proteins were linked to the genomic copy number and spearman's correlation coefficient (*r* with *p*-value) is indicated at the bottom right. (**C**) Correlation between relative peptide and transcriptome abundance in PMC42-ET vs. PMC42-LA per genomic coordinate. Correlation analysis was performed in GraphPad Prism with R^2 value of 0.7361 (*p* < 0.0001).

3.10. The Differences in PMC42 Karyotypes are Reflected in Their Transcriptome and Proteome Ratios

The significant genomic differences in the PMC42 cell lines led us to ask whether the changes in chromosome content mediated changes in transcriptome and proteome that could influence phenotype determination. The relative transcriptome and proteome abundance for a given chromosome across the two PMC42 cell lines were computed from $\log_2$-transformed fold changes of each transcript and proteome. To reduce the noise from the transcriptomic abundance, transcripts that were not expressed or for which normalized values were less than 10 in both the cell lines were discarded. The results show remarkable concordance between the chromosome copy number content and the corresponding transcript and protein abundance from chromosome 3 and chromosome 13 (Figure 9C). PMC42-ET cells have four copies of chromosome 3 on average, whereas PMC42-LA cells have an average of 2 copies, while PMC42-LA cells have three copies of chromosome 13 on average, whereas two copies are present in PMC42-ET. Apart for chromosome 3, transcript and proteome abundance from chromosomes 5, 7, 10 and 16 is also higher in PMC42-ET relative to PMC42-LA, which also corresponds to their higher ploidy distribution in PMC42-ET, except for chromosome 16. Strikingly, the discriminant gene analysis performed for chromosomes 13 and 3 also identified specific genes on chromosome 13 (*DNAJC15, SPG20, SLITRK6* and *DACH1*) that had significant overlap in GSEA with genes down-regulated in TMX2-28 cells (breast cancer) which do not express ESR1 [Gene ID = 2099] compared to the parental MCF-7 cells, which do [72], (p-value 1.6×10^{-5}). Therefore, we hypothesised that the gain in chromosome 13 upregulated the expression of various genes that drives the signalling mechanism of ER in PMC42-LA, causing it to be represented as an upstream regulator in our comparative IPA findings. The results show a high degree of concordance between the relative transcript and proteome abundance across the PMC42 cell lines ($r^2 = 0.736$).

3.11. Bioenergetic Profiles of PMC42 Cells in Comparison with Other Breast Cancer Cell Lines

Since glycolysis was one of the significantly attenuated pathways identified in proteome analysis, we also evaluated the mitochondrial bioenergetic profiles of PMC42 cell lines by measuring their oxygen consumption and glycolysis rates in comparison with the other four breast cancer cell lines: MCF-7, T47D (Luminal), MDA-MB-468 (Basal A) and MDA-MB-231, SUM-159 (Basal B). The extracellular acidification rate (ECAR) and oxygen consumption rate (OCR), as indicators of lactic acid production during glycolysis and mitochondrial respiration during OXPHOS, respectively, were measured (Figure 10). Interestingly, the baseline ECAR status of PMC42-LA was lower than all other cell lines evaluated (Figure 10A). Basal B/mesenchymal MDA-MB-231 and SUM-159 cell lines exhibited higher ECAR as compared to all other cell lines. The higher OCR was seen in luminal MCF-7 and T47D cell lines (Figure 10B).
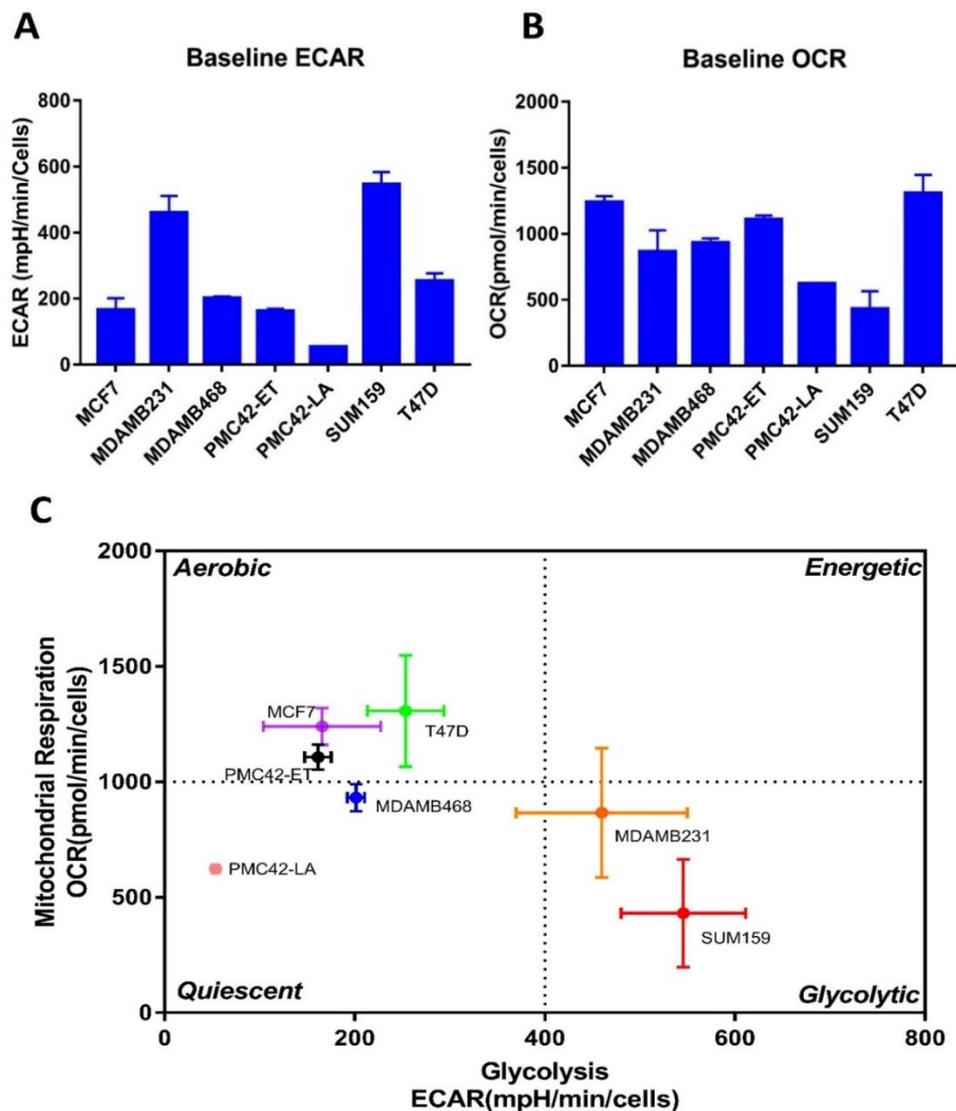
Figure 10. Metabolic profile of PMC42-ET, PMC42-LA and a panel of breast cancer cell lines representative of distinct molecular subtypes. (**A**) Extracellular acidification rate (ECAR) and (**B**) Basal oxygen consumption rate (OCR) measurements. (**C**) OCR: ECAR quadrant showing the bioenergetics phenotype of cell lines using Seahorse analyser (data presented as mean ± s.d., $n = 3$).

4. Discussion

Although the requirement of MET in metastasis is somewhat controversial [21,73–75], the transition of mesenchymally orientated cancer cells to a more epithelial state has been shown to allow cancer cells to survive and seed in distant sites prior to development of a metastatic lesion [18,68,76–78]. The comprehensive integrated analysis of the PMC42 system enhances our understanding of the regulation of molecular events relevant to MET change in the context of breast cancer. In this study, utilization of several omics (exome, transcriptome, proteome) platforms, along with karyotyping and metabolic status, has allowed integrative insights not possible with isolated studies (Figure 11). The PMC42 cell line model system comprises a mesenchymal, parental PMC42-ET cell line and an epithelial derivative PMC42-LA cell line that exhibits profound morphological changes [40], decreased cellular proliferation, distinct karyotype, depletion of TGFBR2 gene, distinct pathways mediated by TNF-alpha signalling, and decreased metabolic bioenergetics.
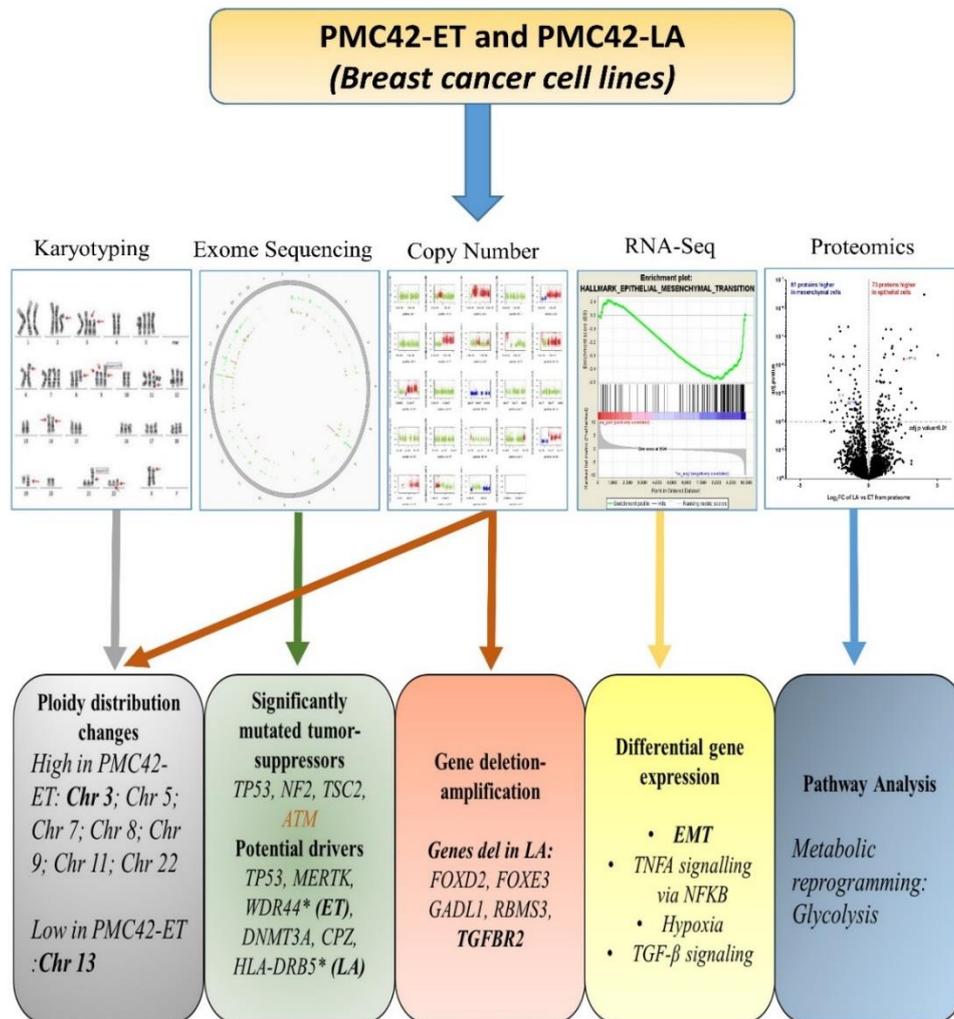
Figure 11. Graphical abstract reflecting the multiple outcomes from the comprehensive and integrative—omics characterization and karyotyping of the PMC42 model system.

Our study identified common canonical driver mutations in TP53, MERTK and DNMT3A in the PMC42 system, as well as a number of unique molecular alterations in the PMC42-ET and PMC42-LA genes that are mutated (Supplementary Tables S5–S8) and have thus potentially impacted the phenotypic heterogeneity seen. Some of the allelic heterogeneous mutations detected in PMC42-LA have also been associated with EMT drivers or markers; we identified novel mutations in *GSN (R397W), WNT1 (T363K), SNAI2 (K188N), ITGA4 (R565W), SOX3 (R22P) and NID2 (G426E)* (Supplementary Table S7). SNAI2-null mice are reported to have retarded epithelial migration rates [79], and SOX3 has been implicated in the malignant behaviour of glioblastoma [80], EMT, and in the promotion of migration and invasion of osteosarcoma cells [81,82]. A N161K mutation in SOX3 was found associated with progression of SCLC along with other mutated SOX members [83]. The implications of genomic variants for these other markers have not been studied or reported in context of EMT. These identified mutations in EMT drivers evaluated at differential gene expression level reflected complete absence of expression of WNT1 and 6-fold downregulation of ITGA4 in PMC42-LA cells.

Comparison of transcriptome and proteome analyses across PMC42 cell lines, combined with GSEA and IPA, were performed to gain insights into the biological processes dysregulated within our model system. In our transcriptome studies, the expression of forkhead box A1 (FOXA1) gene, which acts to control transcription of estrogen receptor-regulated genes and repress the basal-like features of breast cancer cells [84–87], was observed only in PMC42-LA. We observed upregulation in the expression of several specific genes that favour the MET in PMC42-LA, including AGR2 and GRHL2. The most prominent feature of GSEA analysis was the strong down-regulation of EMT in PMC42-LA (*p*-value of < 0.001) (Figure 3). These results indicate that this model system is highly appropriate for studies of EMP [40]. Interestingly, the combined analysis of both transcriptome and proteome across PMC42 cell lines using IPA also indicated that EGFR and ER may be common upstream regulators that are dysregulated between the two cell lines. The ability of EGFR signalling to drive EMT is widely reported in literature [88–90] and is also considered an important driver of MDA-MB-231 invasion leading to formation of brain metastasis [91]. Studies from GSEA and IPA also highlighted an involvement of the TNF-α pathway, mediated by NFKB (*p*-value 0.007), as a major attenuated pathway between PMC42 cell lines. Significant assessment from in vitro studies have highlighted the significance of various cytokines, such as TNF-α [92,93] and growth factors such as EGF and TGF-β in mediating EMT changes, and how these targets can have therapeutic implications in combating EMP [20,93,94].

Furthermore, our recent published work also investigated the dynamic interconversions observed between the transitional epithelial and mesenchymal subpopulations delineated by EpCAM profiling in the predominantly epithelial PMC42-LA breast cancer cells. The subtleties of this transition vary in proportion of epithelial and mesenchymal phenotypes as determined from single cells clonal propagation. Differences observed in the functional attributes of the single cell-derived clones further explains the stochastic nature and the intrinsic cellular plasticity in PMC42-LA. Interestingly, the implementation of whole exome sequencing across the EpCAM-high and -low subpopulation indicates that observed intrinsic phenotypic plasticity in PMC42-LA was not attributable to chromosomal instability [66]. Moreover, the PMC42-LA subline derived from PMC42-ET and remaining phenotypically stable over two decades has maintained distinct karyotype with reduced number of chromosomes and significant ploidy disparity reflected for eight of the chromosomes.

In this pattern of karyotypic differences reflected in the genomic copy number alterations across the two cell lines, we observed that the allele fractions had significantly deviated in the PMC42-LA daughter cell line (*p*-value < 0.0001). PMC42-ET has the higher chromosome number, consistent with a relative increase of its transcriptome and proteome abundance (Figure 9C), a broader range of chromosome numbers, and a higher CIN70 score (Figures 4 and 5). The gain of chromosome 13 and the loss of chromosome 3 in PMC42-LA were the most prominent changes reflected in both the transcriptome and proteome (Figure 9C). Strikingly, the locus of TGFBR2 gene is also on chromosome 3. Notably, we observe a potential link between the deletion affecting the TGF-β receptor gene TGFBR2 in PMC42-LA and its negative TES scores assessed from transcriptome analysis. Moreover, the aberration of TGFBR2 also made PMC42-LA cell line non-responsive to TGF-β1 stimulus for EMT mediated changes (Figure 8C). Thus, changes in specific chromosomes in PMC42-ET, with the acquisition of new mutations and genetic deletion of TGFBR2, may have contributed to the derivation of the MET observed in the PMC42-LA cells. The assessment of the impact of copy number variations at single-gene level across the whole differential mRNA analysis did not yield a significant association (Supplementary Figure S4). Differential gene expression instead of solely based on amplification of gene copies can also be modulated by other factors, such as gene epigenetics or transcription factors. Indeed, Ohshima et al. had associated the gene expression level of various oncogenes with copy number and found that the R value varied between 0.06 and 0.53 across different cancer datasets [95].

CD44 and CD24 are considered putative stem cell markers [96,97] and expression profiling of selected breast cancer cell lines in this study correlates well with previously performed studies on the same cell lines [64,98,99]. Expression of CD24 was equally low or negligible in both the PMC42

cell lines and interestingly, we find higher expression of CD44 in PMC42-LA than anticipated. Many studies have confirmed the CD44^{high}/CD24^{low} profile as a reflection of a mesenchymal state [25,39,100]. However, the epithelially shifted PMC42-LA reflected the increase in CD44^{high}/CD24^{low} profile as compared to the parental PMC42-ET cells. There are other reports of ambiguity regarding functional aspects of these markers [101], suggesting that further investigations of stemness and the regulatory factors in this cell system are required.

Understanding the mechanistic basis of metabolic alterations and their role in tumorigenesis is currently an area of intense research interest [102–106]. Despite glycolysis being one of the significantly attenuated pathways in PMC42-LA cells identified through proteomics analysis, Seahorse experiments indicated decreases in both oxidative metabolism and glycolysis in PMC42-LA cells relative to PMC42-ET. This suggests that the observed metabolic alterations may be attributable to differences in mitochondrial number, rather than proteome differences in metabolic genes required for glycolysis: *ALDH1A3*, *ALDH3A1* and *ALDOC* between cell lines. The analysis of metabolic phenotypes emphasises the importance of using bioenergetics profiles to decipher phenotype assessment, rather than drawing conclusions solely from mutational profiling, RNA-seq and/or proteome data [107,108].

5. Conclusions

In conclusion, the data we have presented herein underscore profound differences in the PMC42 system by allowing for a comprehensive integration of whole exome sequencing, RNA-seq, proteomics and karyotyping data. The identification of novel somatic mutations and indels provide insights into the differences between the two cell lines at genetic scale that potentially drive the phenotypic differences from the genomic level. The loss of *TGFBR2* gene from the derivative PMC42-LA cell line contributes to a reduction in intrinsic TGFB signalling, and these cells are also refractory to TGF-β-induced EMT, however this pathway did not emerge from GSEA analysis. Comparative RNA-seq also demonstrated the PMC42 system to be an authentic MET model. The inter-data relationships illustrate a high degree of concordance between the relative transcript and proteome abundance and chromosome copy number variations across the PMC42 cell lines and identified putative targets to reverse MET. The novel findings provide mechanistic insights into how genomic instability and karyotypic variations have led to acquisition of new autonomous clonal karyotype and phenotype, which is still related to, but distinct from the parental cell line. However, further metabolomics and epigenetics studies might be more compelling to add into the paradigm to determine implications for mesenchymal to epithelial plasticity changes. Overall, this investigation provides an example of the heterogenous changes that may occur in expanding cancer populations and provides evidence for the levels on which these changes occur to affect their phenotypic properties.

References

1. Turashvili, G.; Brogi, E. Tumor Heterogeneity in Breast Cancer. *Front. Med.* **2017**, *4*, 227. [CrossRef] [PubMed]
2. Rivenbark, A.G.; O'Connor, S.M.; Coleman, W.B. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *Am. J. Pathol.* **2013**, *183*, 1113–1124. [CrossRef] [PubMed]
3. Saunus, J.M.; Smart, C.E.; Kutasovic, J.R.; Johnston, R.L.; Kalita-de Croft, P.; Miranda, M.; Rozali, E.N.; Vargas, A.C.; Reid, L.E.; Lorsy, E.; et al. Multidimensional phenotyping of breast cancer cell lines to guide preclinical research. *Breast Cancer Res. Treat.* **2018**, *167*, 289–301. [CrossRef] [PubMed]
4. Keller, P.J.; Lin, A.F.; Arendt, L.M.; Klebba, I.; Jones, A.D.; Rudnick, J.A.; DiMeo, T.A.; Gilmore, H.; Jefferson, D.M.; Graham, R.A.; et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res.* **2010**, *12*, R87. [CrossRef] [PubMed]
5. Hollestelle, A.; Nagel, J.H.; Smid, M.; Lam, S.; Elstrodt, F.; Wasielewski, M.; Ng, S.S.; French, P.J.; Peeters, J.K.; Rozendaal, M.J.; et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res. Treat.* **2010**, *121*, 53–64. [CrossRef] [PubMed]
6. Lehmann, B.D.; Bauer, J.A.; Chen, X.; Sanders, M.E.; Chakravarthy, A.B.; Shyr, Y.; Pietenpol, J.A. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Investig.* **2011**, *121*, 2750–2767. [CrossRef] [PubMed]
7. Neve, R.M.; Chin, K.; Fridlyand, J.; Yeh, J.; Baehner, F.L.; Fevr, T.; Clark, L.; Bayani, N.; Coppé, J.-P.; Tong, F.; et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **2006**, *10*, 515–527. [CrossRef] [PubMed]
8. Prat, A.; Parker, J.S.; Karginova, O.; Fan, C.; Livasy, C.; Herschkowitz, J.I.; He, X.; Perou, C.M. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **2010**, *12*, R68. [CrossRef]
9. Smart, C.E.; Morrison, B.J.; Saunus, J.M.; Vargas, A.C.; Keith, P.; Reid, L.; Wockner, L.; Amiri, M.A.; Sarkar, D.; Simpson, P.T.; et al. In Vitro Analysis of Breast Cancer Cell Line Tumourspheres and Primary Human Breast Epithelia Mammospheres Demonstrates Inter- and Intrasphere Heterogeneity. *PLoS ONE* **2013**, *8*, e64388. [CrossRef]
10. Kao, J.; Salari, K.; Bocanegra, M.; Choi, Y.-L.; Girard, L.; Gandhi, J.; Kwei, K.A.; Hernandez-Boussard, T.; Wang, P.; Gazdar, A.F.; et al. Molecular Profiling of Breast Cancer Cell Lines Defines Relevant Tumor Models and Provides a Resource for Cancer Gene Discovery. *PLoS ONE* **2009**, *4*, e6146. [CrossRef]
11. Daemen, A.; Griffith, O.L.; Heiser, L.M.; Wang, N.J.; Enache, O.M.; Sanborn, Z.; Pepin, F.; Durinck, S.; Korkola, J.E.; Griffith, M.; et al. Modeling precision treatment of breast cancer. *Genome Boil.* **2013**, *14*, R110. [CrossRef]
12. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A.; Kim, S.; Wilson, C.J.; Lehár, J.; Kryukov, G.V.; Sonkin, D.; et al. Addendum: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *492*, 290. [CrossRef]
13. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell* **2017**, *171*, 1437–1452. [CrossRef]
14. Cope, L.M.; Fackler, M.J.; Lopez-Bujanda, Z.; Wolff, A.C.; Visvanathan, K.; Gray, J.W.; Sukumar, S.; Umbricht, C.B. Do Breast Cancer Cell Lines Provide a Relevant Model of the Patient Tumor Methylome? *PLoS ONE* **2014**, *9*, e105545. [CrossRef]
15. Jiang, G.; Zhang, S.; Yazdanparast, A.; Li, M.; Pawar, A.V.; Liu, Y.; Inavolu, S.M.; Cheng, L. Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. *BMC Genom.* **2016**, *17*, 1079. [CrossRef]
16. Heiser, L.M.; Sadanandam, A.; Kuo, W.L.; Benz, S.C.; Goldstein, T.C.; Ng, S.; Gibb, W.J.; Wang, N.J.; Ziyad, S.; Tong, F.; et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2724–2729. [CrossRef]
17. Hugo, H.; Ackland, M.L.; Blick, T.; Lawrence, M.G.; Clements, J.A.; Williams, E.D.; Thompson, E.W. Epithelial—mesenchymal and mesenchymal—epithelial transitions in carcinoma progression. *J. Cell. Physiol.* **2007**, *213*, 374–383. [CrossRef]
18. Nieto, M.A.; Huang, R.Y.J.; Jackson, R.A.; Thiery, J.P. EMT: 2016. *Cell* **2016**, *166*, 21–45. [CrossRef]

19. Jolly, M.K.; Tripathi, S.C.; Jia, D.; Mooney, S.M.; Celiktas, M.; Hanash, S.M.; Mani, S.A.; Pienta, K.J.; Ben-Jacob, E.; Levine, H. Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **2016**, *7*, 27067–27084. [CrossRef]

20. Bhatia, S.; Monkman, J.; Toh, A.K.L.; Nagaraj, S.H.; Thompson, E.W. Targeting epithelial–mesenchymal plasticity in cancer: clinical and preclinical advances in therapy and monitoring. *Biochem. J.* **2017**, *474*, 3269–3306. [CrossRef]

21. Fischer, K.R.; Durrans, A.; Lee, S.; Sheng, J.; Li, F.; Wong, S.T.; Choi, H.; El Rayes, T.; Ryu, S.; Troeger, J.; et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **2015**, *527*, 472–476. [CrossRef]

22. Zheng, X.; Carstens, J.L.; Kim, J.; Scheible, M.; Kaye, J.; Sugimoto, H.; Wu, C.C.; LeBleu, V.S.; Kalluri, R. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **2015**, *527*, 525–530. [CrossRef]

23. Redfern, A.D.; Spalding, L.J.; Thompson, E.W. The Kraken Wakes: induced EMT as a driver of tumour aggression and poor outcome. *Clin. Exp. Metastasis* **2018**, *35*, 285–308. [CrossRef]

24. Blick, T.; Widodo, E.; Hugo, H.; Waltham, M.; Lenburg, M.E.; Neve, R.M.; Thompson, E.W. Epithelial mesenchymal transition traits in human breast cancer cell lines. *Clin. Exp. Metastasis* **2008**, *25*, 629–642. [CrossRef]

25. Blick, T.; Hugo, H.; Widodo, E.; Waltham, M.; Pinto, C.; Mani, S.A.; Weinberg, R.A.; Neve, R.M.; Lenburg, M.E.; Thompson, E.W. Epithelial Mesenchymal Transition Traits in Human Breast Cancer Cell Lines Parallel the $CD_{44}^{hi}/CD_{24}^{lo/-}$ Stem Cell Phenotype in Human Breast Cancer. *J. Mammary Gland. Boil. Neoplasia* **2010**, *15*, 235–252. [CrossRef]

26. Tan, T.Z.; Miow, Q.H.; Miki, Y.; Noda, T.; Mori, S.; Huang, R.Y.-J.; Thiery, J.P. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **2014**, *6*, 1279–1293. [CrossRef]

27. Foroutan, M.; Cursons, J.; Hediyeh-Zadeh, S.; Thompson, E.W.; Davis, M.J. A Transcriptional Program for Detecting TGFbeta-Induced EMT in Cancer. *Mol. Cancer Res.* **2017**, *15*, 619–631. [CrossRef]

28. Foroutan, M.; Bhuva, D.D.; Lyu, R.; Horan, K.; Cursons, J.; Davis, M.J. Single sample scoring of molecular phenotypes. *BMC Bioinform.* **2018**, *19*, 404. [CrossRef]

29. George, J.T.; Jolly, M.K.; Xu, S.; Somarelli, J.A.; Levine, H. Survival outcomes in cancer patients predicted by a partial EMT gene expression scoring metric. *Cancer Res.* **2017**, *77*, 6415–6428. [CrossRef]

30. Cursons, J.; Leuchowius, K.-J.; Waltham, M.; Tomaskovic-Crook, E.; Foroutan, M.; Bracken, C.P.; Redfern, A.; Crampin, E.J.; Street, I.; Davis, M.J.; et al. Stimulus-dependent differences in signalling regulate epithelial-mesenchymal plasticity and change the effects of drugs in breast cancer cell lines. *Cell Commun. Signal.* **2015**, *13*, 26. [CrossRef]

31. Buonato, J.M.; Lan, I.S.; Lazzara, M.J. EGF augments TGFβ-induced epithelial–mesenchymal transition by promoting SHP2 binding to GAB1. *J. Cell Sci.* **2015**, *128*, 3898–3909. [CrossRef]

32. Hugo, H.J.; Wafai, R.; Blick, T.; Thompson, E.W.; Newgreen, D.F. Staurosporine augments EGF-mediated EMT in PMC42-LA cells through actin depolymerisation, focal contact size reduction and Snail1 induction—A model for cross-modulation. *BMC Cancer* **2009**, *9*, 235. [CrossRef]

33. Kurimoto, R.; Iwasawa, S.; Ebata, T.; Ishiwata, T.; Sekine, I.; Tada, Y.; Tatsumi, K.; Koide, S.; Iwama, A.; Takiguchi, Y. Drug resistance originating from a TGF-beta/FGF-2-driven epithelial-to-mesenchymal transition and its reversion in human lung adenocarcinoma cell lines harboring an EGFR mutation. *Int. J. Oncol.* **2016**, *48*, 1825–1836. [CrossRef]

34. Tiwari, N.; Tiwari, V.K.; Waldmeier, L.; Balwierz, P.J.; Arnold, P.; Pachkov, M.; Meyer-Schaller, N.; Schübeler, D.; Van Nimwegen, E.; Christofori, G. Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. *Cancer Cell* **2013**, *23*, 768–783. [CrossRef]

35. Yamamoto, M.; Sakane, K.; Tominaga, K.; Gotoh, N.; Niwa, T.; Kikuchi, Y.; Tada, K.; Goshima, N.; Semba, K.; Inoue, J. Intratumoral bidirectional transitions between epithelial and mesenchymal cells in triple-negative breast cancer. *Cancer Sci.* **2017**, *108*, 1210–1222. [CrossRef]

36. Taube, J.H.; Herschkowitz, J.I.; Komurov, K.; Zhou, A.Y.; Gupta, S.; Yang, J.; Hartwell, K.; Onder, T.T.; Gupta, P.B.; Evans, K.W.; et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15449–15454. [CrossRef]

37. Morel, A.-P.; Lièvre, M.; Thomas, C.; Hinkal, G.; Ansieau, S.; Puisieux, A. Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. *PLoS ONE* **2008**, *3*, e2888. [CrossRef]

38. Briem, E.; Ingthorsson, S.; Traustadottir, G.A.; Hilmarsdottir, B.; Gudjonsson, T. Application of the D492 Cell Lines to Explore Breast Morphogenesis, EMT and Cancer Progression in 3D Culture. *J. Mammary Gland. Boil. Neoplasia* **2019**, *24*, 139–147. [CrossRef]

39. Mani, S.A.; Guo, W.; Liao, M.-J.; Eaton, E.N.; Ayyanan, A.; Zhou, A.Y.; Brooks, M.; Reinhard, F.; Zhang, C.C.; Shipitsin, M.; et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **2008**, *133*, 704–715. [CrossRef]

40. Hugo, H.J.; Kokkinos, M.I.; Blick, T.; Ackland, M.L.; Thompson, E.W.; Newgreen, D.F. Defining the E-Cadherin Repressor Interactome in Epithelial-Mesenchymal Transition: The PMC42 Model as a Case Study. *Cells Tissues Organs* **2011**, *193*, 23–40. [CrossRef]

41. Whitehead, R.H.; Bertoncello, I.; Webber, L.M.; Pedersen, J.S. A new human breast carcinoma cell line (PMC42) with stem cell characteristics. I. Morphologic characterization. *J. Natl. Cancer Inst.* **1983**, *70*, 649–661.

42. Whitehead, R.H.; Webber, L.M.; Bertoncello, I.; Vitali, A.A.; Monaghan, P. A New Human Breast Carcinoma Cell Line (PMC42) With Stem Cell Characteristics. II. Characterization of Cells Growing as Organoids. *J. Natl. Cancer Inst.* **1983**, *71*, 1193–1203.

43. Whitehead, R.H.; Quirk, S.J.; Vitali, A.A.; Funder, J.W.; Sutherland, R.L.; Murphy, L.C. A new human breast carcinoma cell line (PMC42) with stem cell characteristics. III. Hormone receptor status and responsiveness. *J. Natl. Cancer Inst.* **1984**, *73*, 643–648.

44. Ackland, M.L.; Newgreen, D.F.; Fridman, M.; Waltham, M.C.; Arvanitis, A.; Minichiello, J.; Price, J.T.; Thompson, E.W. Epidermal Growth Factor-Induced Epithelio-Mesenchymal Transition in Human Breast Carcinoma Cells. *Lab. Investig.* **2003**, *83*, 435–448. [CrossRef]

45. Ackland, M.; Michalczyk, A.; Whitehead, R. PMC42, A Novel Model for the Differentiated Human Breast. *Exp. Cell Res.* **2001**, *263*, 14–22. [CrossRef]

46. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]

47. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]

48. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]

49. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef]

50. ANNOVAR. Available online: http://openbioinformatics.org/annovar/ (accessed on 30 January 2019).

51. Boeva, V.; Popova, T.; Bleakley, K.; Chiche, P.; Cappo, J.; Schleiermacher, G.; Janoueix-Lerosey, I.; Delattre, O.; Barillot, E. Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **2012**, *28*, 423–425. [CrossRef]

52. Douville, C.; Carter, H.; Kim, R.; Niknafs, N.; Diekhans, M.; Stenson, P.D.; Cooper, D.N.; Ryan, M.; Karchin, R. CRAVAT: cancer-related analysis of variants toolkit. *Bioinformatics* **2013**, *29*, 647–648. [CrossRef]

53. Carter, H.; Samayoa, J.; Hruban, R.H.; Karchin, R. Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). *Cancer Boil. Ther.* **2010**, *10*, 582–587. [CrossRef]

54. Carter, H.; Chen, S.; Isik, L.; Tyekucheva, S.; Velculescu, V.E.; Kinzler, K.W.; Vogelstein, B.; Karchin, R. Cancer-specific High-throughput Annotation of Somatic Mutations: computational prediction of driver missense mutations. *Cancer Res.* **2009**, *69*, 6660–6667. [CrossRef]

55. Carter, H.; Douville, C.; Stenson, P.D.; Cooper, D.N.; Karchin, R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. *BMC Genom.* **2013**, *14*, S3. [CrossRef]

56. Morpheus. Available online: https://software.broadinstitute.org/morpheus/ (accessed on 25 Febrary 2019).

57. Mar, J.C.; Kimura, Y.; Schroder, K.; Irvine, K.M.; Hayashizaki, Y.; Suzuki, H.; Hume, D.; Quackenbush, J. Data-driven normalization strategies for high-throughput quantitative RT-PCR. *BMC Bioinform.* **2009**, *10*, 110. [CrossRef]

58. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [CrossRef]

59. GSEA. Available online: http://www.broad.mit.edu/gsea (accessed on 24 April 2019).

60. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [CrossRef]

61. UniProt. Available online: https://www.uniprot.org/ (accessed on 25 November 2018).

62. PeakView Software. Available online: https://sciex.com/products/software/peakview-software (accessed on 25 November 2018).

63. Clough, T.; Thaminy, S.; Ragg, S.; Aebersold, R.; Vitek, O. Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. *BMC Bioinform.* **2012**, *13*, S6. [CrossRef]

64. Ricardo, S.; Vieira, A.F.; Gerhard, R.; Leitão, D.; Pinto, R.; Cameselle-Teijeiro, J.F.; Milanezi, F.; Schmitt, F.; Paredes, J. Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. *J. Clin. Pathol.* **2011**, *64*, 937–946. [CrossRef]

65. Matalkah, F.; Martin, E.; Zhao, H.; Agazie, Y.M. SHP2 acts both upstream and downstream of multiple receptor tyrosine kinases to promote basal-like and triple-negative breast cancer. *Breast Cancer Res.* **2016**, *18*, 25. [CrossRef]

66. Bhatia, S.; Monkman, J.; Blick, T.; Pinto, C.; Waltham, M.; Nagaraj, S.H.; Thompson, E.W. Interrogation of Phenotypic Plasticity between Epithelial and Mesenchymal States in Breast Cancer. *J. Clin. Med.* **2019**, *8*, 893. [CrossRef]

67. Carter, S.L.; Eklund, A.C.; Kohane, I.S.; Harris, L.N.; Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **2006**, *38*, 1043–1048. [CrossRef]

68. Gunasinghe, N.P.A.D.; Wells, A.; Thompson, E.W.; Hugo, H.J. Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. *Cancer Metastasis Rev.* **2012**, *31*, 469–478. [CrossRef]

69. Tsubakihara, Y.; Moustakas, A. Epithelial–mesenchymal Transition and Metastasis under the Control of Transforming Growth Factor beta. *Int. J. Mol. Sci.* **2018**, *19*, 3672. [CrossRef]

70. Pino, M.S.; Kikuchi, H.; Zeng, M.; Herraiz, M.T.; Sperduti, I.; Berger, D.; Park, D.Y.; Iafrate, A.J.; Zukerberg, L.R.; Chung, D.C. Epithelial to mesenchymal transition is impaired in colon cancer cells with microsatellite instability. *Gastroenterology* **2010**, *138*, 1406–1417. [CrossRef]

71. Sommerova, L.; Ondrouskova, E.; Vojtesek, B.; Hrstka, R. Suppression of AGR2 in a TGF-β-induced Smad regulatory pathway mediates epithelial-mesenchymal transition. *BMC Cancer* **2017**, *17*, 546. [CrossRef]

72. Gozgit, J.M.; Pentecost, B.T.; Marconi, S.A.; Ricketts-Loriaux, R.S.J.; Otis, C.N.; Arcaro, K.F. PLD1 is overexpressed in an ER-negative MCF-7 cell line variant and a subset of phospho-Akt-negative breast carcinomas. *Br. J. Cancer* **2007**, *97*, 809–817. [CrossRef]

73. Ye, X.; Brabletz, T.; Kang, Y.; Longmore, G.D.; Nieto, M.A.; Stanger, B.Z.; Yang, J.; Weinberg, R.A. Upholding a role for EMT in breast cancer metastasis. *Nature* **2017**, *547*, E1–E3. [CrossRef]

74. Fischer, K.R.; Altorki, N.K.; Mittal, V.; Gao, D. Fischer et al. reply. *Nature* **2017**, *547*, E5. [CrossRef]

75. Brabletz, T.; Kalluri, R. EMT in cancer. *Nat. Rev. Cancer* **2018**, *18*, 128–134. [CrossRef]

76. Beerling, E.; Seinstra, D.; De Wit, E.; Kester, L.; Van Der Velden, D.; Maynard, C.; Schäfer, R.; Van Diest, P.; Voest, E.; Van Oudenaarden, A.; et al. Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep.* **2016**, *14*, 2281–2288. [CrossRef]

77. Brabletz, T. To differentiate or not—Routes towards metastasis. *Nat. Rev. Cancer* **2012**, *12*, 425–436. [CrossRef]

78. van Denderen, B.J.; Thompson, E.W. Cancer: The to and fro of tumour spread. *Nature* **2013**, *493*, 487–488. [CrossRef]

79. Savagner, P.; Kusewitt, D.F.; Carver, E.A.; Magnino, F.; Choi, C.; Gridley, T.; Hudson, L.G. Developmental transcription factor slug is required for effective re-epithelialization by adult keratinocytes. *J. Cell. Physiol.* **2005**, *202*, 858–866. [CrossRef]

80. Vicentic, J.M.; Drakulic, D.; Garcia, I.; Vukovic, V.; Aldaz, P.; Puskas, N.; Nikolic, I.; Tasic, G.; Raicevic, S.; Garros-Regulez, L.; et al. SOX3 can promote the malignant behavior of glioblastoma cells. *Cell Oncol.* **2019**, *42*, 41–54. [CrossRef]

81. Guo, Y.; Yin, J.; Tang, M.; Yu, X. Downregulation of SOX3 leads to the inhibition of the proliferation, migration and invasion of osteosarcoma cells. *Int. J. Oncol.* **2018**, *52*, 1277–1284. [CrossRef]

82. Qiu, M.; Chen, D.; Shen, C.; Shen, J.; Zhao, H.; He, Y. Sex-determining region Y-box protein 3 induces epithelial-mesenchymal transition in osteosarcoma cells via transcriptional activation of Snail1. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 46. [CrossRef]

83. Rudin, C.M.; Durinck, S.; Stawiski, E.W.; Poirier, J.T.; Modrusan, Z.; Shames, D.S.; Bergbower, E.A.; Guan, Y.; Shin, J.; Guillory, J.; et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat. Genet.* **2012**, *44*, 1111–1116. [CrossRef]

84. Badve, S.; Turbin, D.; Thorat, M.A.; Morimiya, A.; Nielsen, T.O.; Perou, C.M.; Dunn, S.; Huntsman, D.G.; Nakshatri, H. FOXA1 Expression in Breast Cancer Correlation with Luminal Subtype A and Survival. *Clin. Cancer Res.* **2007**, *13*, 4415–4421. [CrossRef]

85. Rangel, N.; Fortunati, N.; Osella-Abate, S.; Annaratone, L.; Isella, C.; Catalano, M.G.; Rinella, L.; Metovic, J.; Boldorini, R.; Balmativola, D.; et al. FOXA1 and AR in invasive breast cancer: new findings on their co-expression and impact on prognosis in ER-positive patients. *BMC Cancer* **2018**, *18*, 703. [CrossRef]

86. Fu, X.; Jeselsohn, R.; Pereira, R.; Hollingsworth, E.F.; Creighton, C.J.; Li, F.; Shea, M.; Nardone, A.; De Angelis, C.; Heiser, L.M.; et al. FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome and IL-8 expression in ER-positive breast cancer. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6600–E6609. [CrossRef]

87. Bernardo, G.M.; Bebek, G.; Ginther, C.L.; Sizemore, S.T.; Lozada, K.L.; Miedler, J.D.; Anderson, L.A.; Godwin, A.K.; Abdul-Karim, F.W.; Slamon, D.J.; et al. FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene* **2013**, *32*, 554–563. [CrossRef]

88. Al Moustafa, A.-E.; Ala-Eddin, A.M. EGF-receptor signaling and epithelial-mesenchymal transition in human carcinomas. *Front. Biosci.* **2012**, *4*, 671–684. [CrossRef]

89. Lo, H.-W.; Hsu, S.-C.; Xia, W.; Cao, X.; Shih, J.-Y.; Wei, Y.; Abbruzzese, J.L.; Hortobagyi, G.N.; Hung, M.-C. Epidermal Growth Factor Receptor Cooperates with Signal Transducer and Activator of Transcription 3 to Induce Epithelial-Mesenchymal Transition in Cancer Cells via Up-regulation of TWIST Gene Expression. *Cancer Res.* **2007**, *67*, 9066–9076. [CrossRef]

90. Mizumoto, A.; Yamamoto, K.; Nakayama, Y.; Takara, K.; Nakagawa, T.; Hirano, T.; Hirai, M. Induction of epithelial–mesenchymal transition via activation of epidermal growth factor receptor contributes to sunitinib resistance in human renal cell carcinoma cell lines. *J. Pharmacol. Exp. Ther.* **2015**, *355*, 152–158. [CrossRef]

91. Nie, F.; Yang, J.; Wen, S.; An, Y.-L.; Ding, J.; Ju, S.-H.; Zhao, Z.; Chen, H.-J.; Peng, X.-G.; Wong, S.T.C.; et al. Involvement of epidermal growth factor receptor overexpression in the promotion of breast cancer brain metastasis. *Cancer* **2012**, *118*, 5198–5209. [CrossRef]

92. Tang, D.; Tao, D.; Fang, Y.; Deng, C.; Xu, Q.; Zhou, J. TNF-Alpha Promotes Invasion and Metastasis via NF-Kappa B Pathway in Oral Squamous Cell Carcinoma. *Med Sci. Monit. Basic Res.* **2017**, *23*, 141–149. [CrossRef]

93. Li, C.W.; Xia, W.; Huo, L.; Lim, S.O.; Wu, Y.; Hsu, J.L.; Chao, C.H.; Yamaguchi, H.; Yang, N.K.; Ding, Q.; et al. Epithelial–mesenchymal transition induced by TNF-alpha requires NF-kappaB-mediated transcriptional upregulation of Twist1. *Cancer Res.* **2012**, *72*, 1290–1300. [CrossRef]

94. Marcucci, F.; Stassi, G.; De Maria, R. Epithelial–mesenchymal transition: a new target in anticancer drug discovery. *Nat. Rev. Drug Discov.* **2016**, *15*, 311–325. [CrossRef]

95. Ohshima, K.; Hatakeyama, K.; Nagashima, T.; Watanabe, Y.; Kanto, K.; Doi, Y.; Ide, T.; Shimoda, Y.; Tanabe, T.; Ohnami, S.; et al. Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Sci. Rep.* **2017**, *7*, 1546. [CrossRef]

96. Hurt, E.M.; Kawasaki, B.T.; Klarmann, G.J.; Thomas, S.B.; Farrar, W.L. CD44+ CD24− prostate cells are early cancer progenitor/stem cells that provide a model for patients with poor prognosis. *Br. J. Cancer* **2008**, *98*, 756–765. [CrossRef]

97. Klonisch, T.; Wiechec, E.; Hombach-Klonisch, S.; Ande, S.R.; Wesselborg, S.; Schulze-Osthoff, K.; Los, M. Cancer stem cell markers in common cancers—Therapeutic implications. *Trends Mol. Med.* **2008**, *14*, 450–460. [CrossRef]

98. Li, W.; Ma, H.; Zhang, J.; Zhu, L.; Wang, C.; Yang, Y. Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. *Sci. Rep.* **2017**, *7*, 13856. [CrossRef]

99. Fillmore, C.M.; Kuperwasser, C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res.* **2008**, *10*, R25. [CrossRef]

100. Bane, A.; Viloria-Petit, A.; Pinnaduwage, D.; Mulligan, A.M.; O'Malley, F.P.; Andrulis, I.L. Clinical–pathologic significance of cancer stem cell marker expression in familial breast cancers. *Breast Cancer Res. Treat.* **2013**, *140*, 195–205. [CrossRef]

101. Jaggupilli, A.; Elkord, E. Significance of CD44 and CD24 as Cancer Stem Cell Markers: An Enduring Ambiguity. *Clin. Dev. Immunol.* **2012**, *2012*, 1–11. [CrossRef]

102. Jones, N.P.; Schulze, A. Targeting cancer metabolism—Aiming at a tumour's sweet-spot. *Drug Discov. Today* **2012**, *17*, 232–241. [CrossRef]

103. Heiden, M.G.V. Targeting cancer metabolism: a therapeutic window opens. *Nat. Rev. Drug Discov.* **2011**, *10*, 671–684. [CrossRef]

104. Meijer, T.W.; Peeters, W.J.; Dubois, L.J.; Van Gisbergen, M.W.; Biemans, R.; Venhuizen, J.-H.; Span, P.N.; Bussink, J. Targeting glucose and glutamine metabolism combined with radiation therapy in non-small cell lung cancer. *Lung Cancer* **2018**, *126*, 32–40. [CrossRef]

105. Chao, T.-K.; Huang, T.-S.; Liao, Y.-P.; Huang, R.-L.; Su, P.-H.; Shen, H.-Y.; Lai, H.-C.; Wang, Y.-C. Pyruvate kinase M2 is a poor prognostic marker of and a therapeutic target in ovarian cancer. *PLoS ONE* **2017**, *12*, e0182166. [CrossRef]

106. Shackelford, D.B.; Abt, E.; Gerken, L.; Vasquez, D.S.; Seki, A.; Leblanc, M.; Wei, L.; Fishbein, M.C.; Czernin, J.; Mischel, P.S.; et al. LKB1 inactivation dictates therapeutic response of non-small cell lung cancer to the metabolism drug phenformin. *Cancer Cell* **2013**, *23*, 143–158. [CrossRef]

107. Hardie, R.-A.; Van Dam, E.; Cowley, M.; Han, T.-L.; Balaban, S.; Pajic, M.; Pinese, M.; Iconomou, M.; Shearer, R.F.; McKenna, J.; et al. Mitochondrial mutations and metabolic adaptation in pancreatic cancer. *Cancer Metab.* **2017**, *5*, 7. [CrossRef]

108. Pelicano, H.; Zhang, W.; Liu, J.; Hammoudi, N.; Dai, J.; Xu, R.-H.; Pusztai, L.; Huang, P. Mitochondrial dysfunction in some triple-negative breast cancer cell lines: role of mTOR pathway and therapeutic potential. *Breast Cancer Res.* **2014**, *16*, 434. [CrossRef]