**Materials and Methods:**

1. **Cell Lines and culture**

The cell lines MDA-MB-231, MCF7 and HCC1937 were obtained from the American Type Culture Collection (ATCC- Manassas, VA). MDA-MB-468 and ZR-75-1 were obtained from NCCS (Pune, India) where cell authentication was performed using STR profiling. MDA-MB-231 and MDA-MB-468 were maintained in L-15 (Leibovitz) medium (Sigma-Aldrich), MCF7 in DMEM-Hi Glucose medium (Sigma-Aldrich) HCC1937 and ZR-75-1 in RPMI 1640 media (Gibco), HEPES buffered and supplemented with 10% (v/v) heat inactivated Fetal Bovine Serum (Himedia) and 100 U/ml penicillin and streptomycin (Gibco). All cells were maintained in a humidified incubator with 5% CO2 at 37°C except for MDA-MB-231 and MDA-MB-468 that were maintained with 0% CO2. For all experimental assays using cell lines, a passage number below 20 was used and all cell lines were subjected to frequent recharacterization by immunophenotyping and testing of mycoplasma.

1. **Western blot**

Cells (15\*104) were seeded in 12-well plates and after 24-48 hours, lysed and the protein expression was assayed as reported previously. The antibodies used were anti-E-cadherin (Abcam-EP700Y) and anti-Vimentin (BioGenex) and incubation was performed overnight at 4°C at specific dilutions- 1:5000 and 1:100 respectively. The membrane was then washed and incubated with goat anti-mouse/rabbit IgG conjugated to horseradish peroxidase (Jackson Immunoresearch). Immunoreactive bands were visualized using enhanced chemiluminescence. Densitometric analysis was performed using quantity one software (Bio-rad) as reported previously [1].

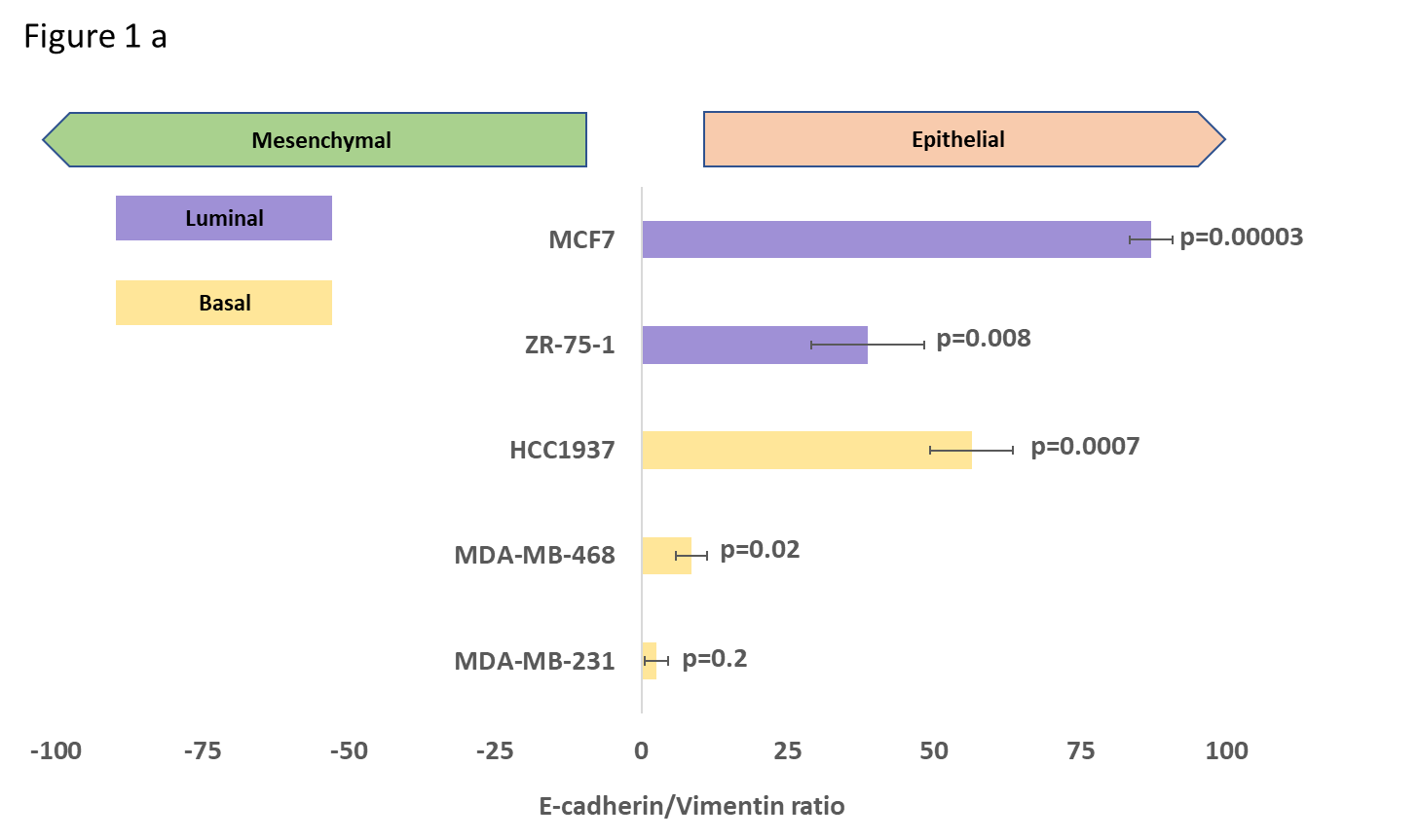
1. **Dual immunofluorescence**

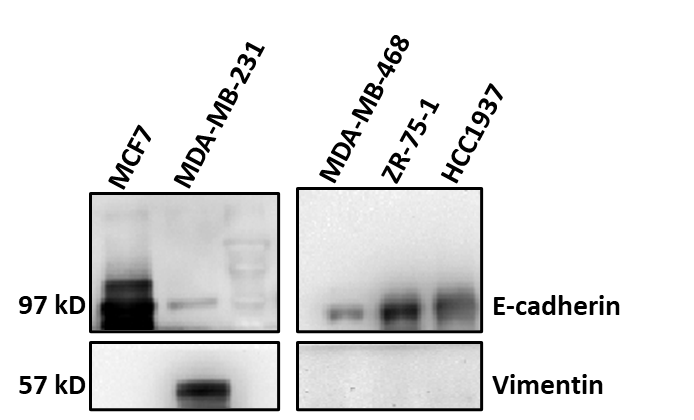
Cells (1\*104) were seeded on poly-L-lysine coated coverslips. Immunofluorescence was performed as reported previously [1] by incubating cells in primary antibodies- anti-E-cadherin (Abcam-EP700Y) and anti-Vimentin (BioGenex) overnight at 4°C at specific dilutions- 1:500 and 1:25 respectively. This was followed by labelling with specific secondary antibodies - Alexa Fluor® 488 Chicken Anti-Mouse IgG (H+L) for Anti-Vimentin and Alexa Fluor 568 Donkey Anti-Rabbit IgG for anti-E-cadherin for 1 h at room temperature. The slide was then mounted on gold antifade reagent with DAPI and examined under a fluorescent microscope (Olympus BX51).

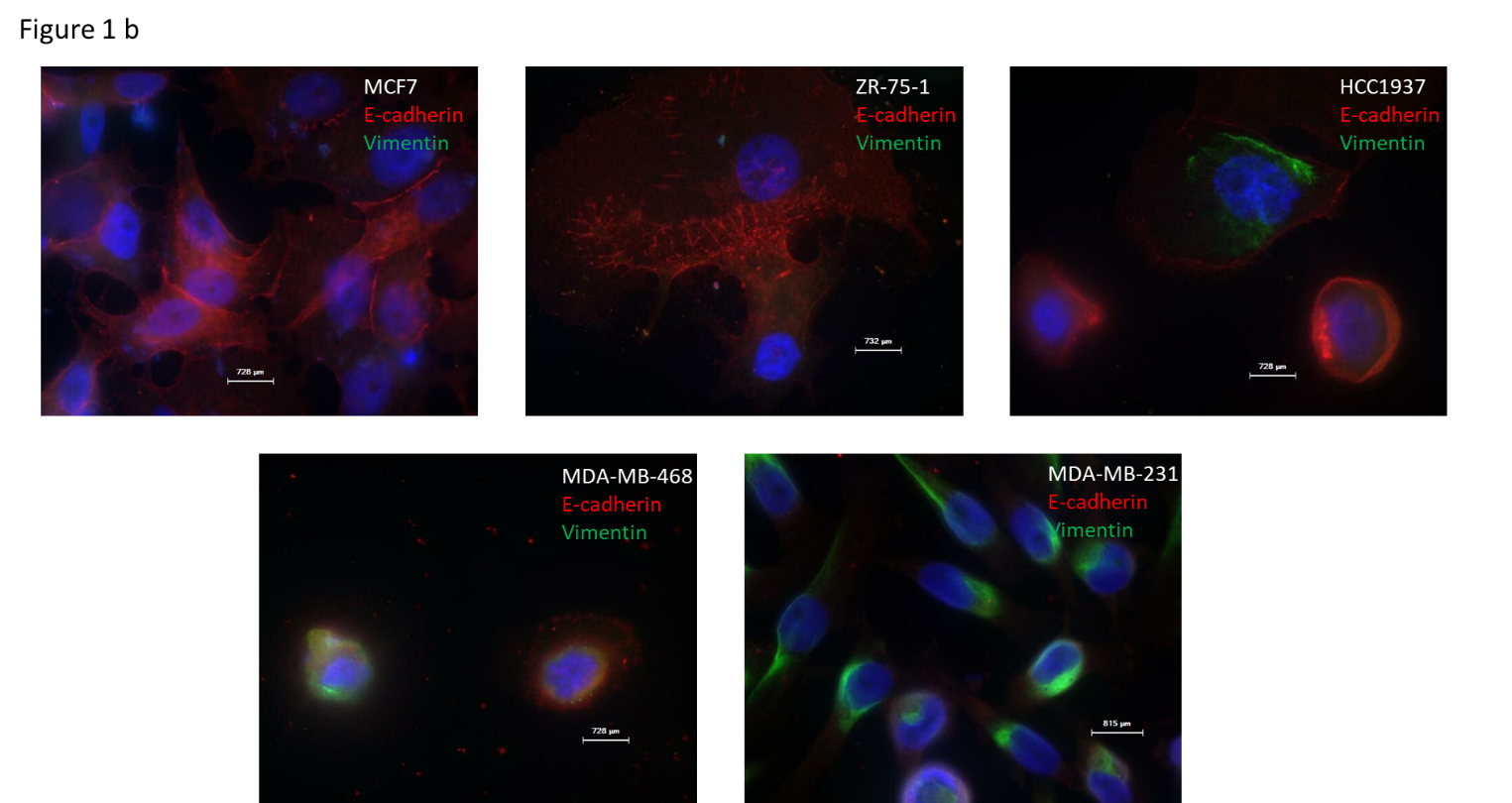
1. **mRNA expression analysis using quantitative PCR**

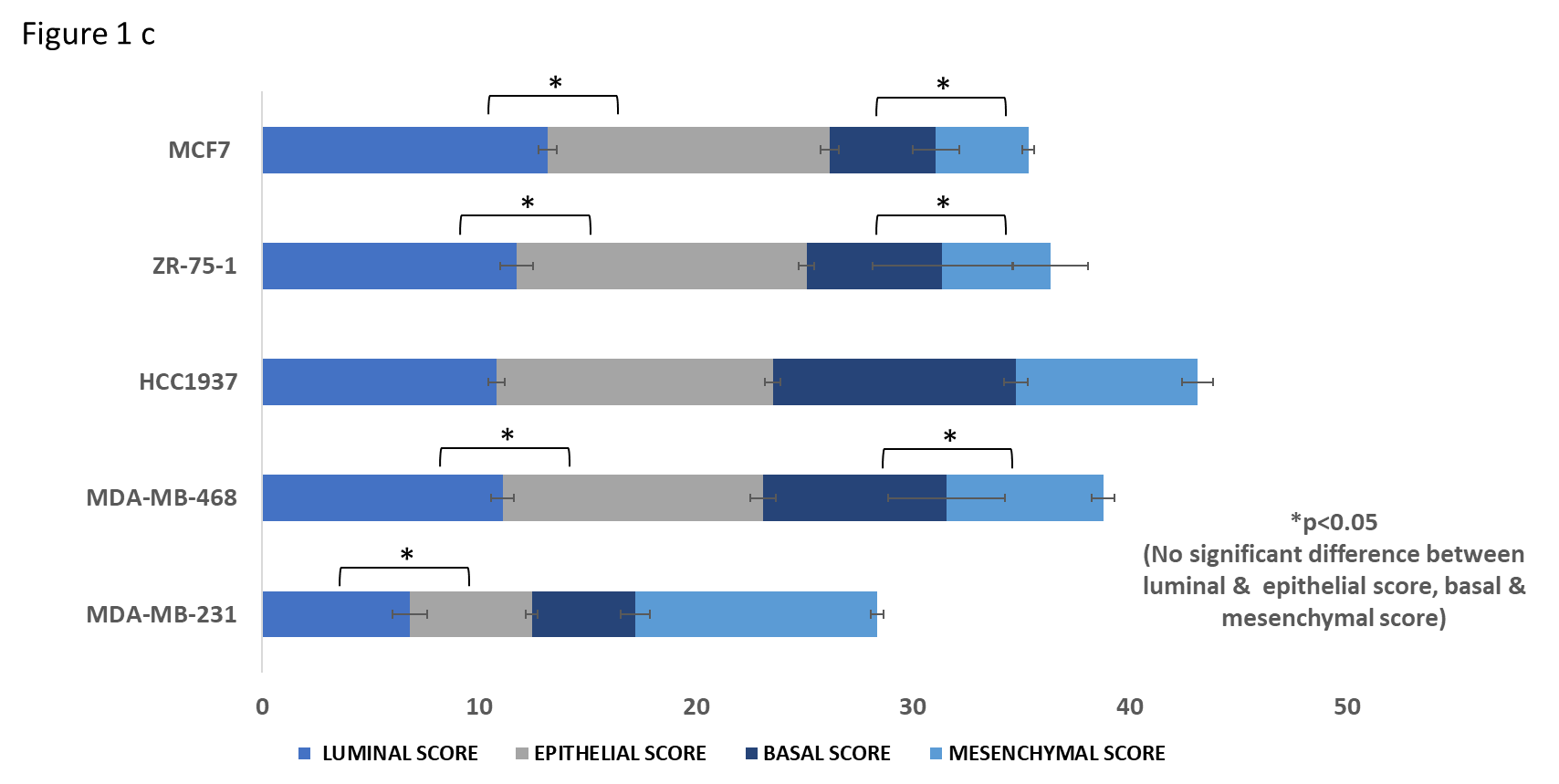
Extraction of RNA, cDNA synthesis, and q-PCR experiments were performed on cell line lysates as reported previously [1, 2]. Total RNA was extracted using TRI Reagent (Sigma-Aldrich) according to manufacturer's instructions. 500ng of total RNA was then reverse transcribed into cDNA using the ABI high-capacity cDNA archive kit (ABI) as per the manufacturer's protocol. The primer sequences for the genes tested are given in Table 1. Expression level of housekeeping genes were determined using a panel of 3 reference genes – *ACTB, RPLP0 and PUM1*. A total reaction volume of 10 μl containing 1 ng of cDNA template per qRT-PCR reaction was done in duplicate using SYBR Green chemistry and run on LightCycler 480 II (Roche Diagnostics).

**Figures:**

****

****

****

****

**Figure legend:**

**Figure 1. Phenotypic characterisation of luminal and basal cell lines-** (a) E-cadherin/Vimentin ratio as evaluated by western blot analysis (b) Immunofluorescence analysis of E-cadherin and Vimentin expression (c) Gene expression analysis for computation of luminal, epithelial, basal and mesenchymal scores in the cell lines- MCF7, ZR-75-1, HCC1937, MDA-MB-468 and MDA-MB-231.

**Results:**

Cell lines representing the various subtypes; MCF7 and ZR-75-1 representing luminal breast cancer subtype, MDA-MB-231, MDA-MB-468 and HCC1937 representing the basal cell lines were used for the in-vitro experiments. E-cadherin and Vimentin protein expression was quantified by western blot and a ratio (E-cadherin/Vimentin) was computed (Figure 1 a). The ratio was highest in the luminal cell lines -MCF7 (87.07; p=0.00003) followed by ZR-75-1 (38.69; p=0.008) as they expressed high levels of E-cadherin as confirmed by immunofluorescence also (Figure 1 b). Among the basal cell lines HCC1937 expressed a higher ratio (56.4; p=0.0007) owing to higher levels of E-cadherin expression. However, MDA-MB-231 and MDA-MB-468 expressed a lower ratio- 8.55 (p=0.02) and 2.57 (p=0.2) respectively. MDA-MB-231 expressed the lowest ratio indicative of equivalence in terms of E-cadherin and Vimentin expression.

Gene expression analysis was performed using multiple genes indicative of the luminal, epithelial, basal and mesenchymal phenotype as described in Table 1. Luminal, epithelial, basal and mesenchymal score was computed using average gene expression levels of the genes ascribed to each phenotype. The luminal cell lines - MCF7 and ZR-75-1 expressed near equivalence in terms of a luminal and epithelial score indicative of a good correlation between the two phenotypes. Among the basal cell lines, HCC1937 displayed the highest basal score but expressed higher levels of epithelial score as well. MDA-MB-231 expressed the highest mesenchymal score and the least epithelial score (Figure c, Table 2).

Table 1: List of primer sequences used for gene expression analysis for computation of luminal, epithelial, basal and mesenchymal scores.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl no.** | **Gene** | **Utility** | **Primer sequence** |
| 1 | *ACTB* | Housekeeping | 5’-TTCCTGGGCATGGAGTC-3’ |
|  |  |  | 3’-CAGGTCTTTGCGGATGTC-5’ |
| 2 | *RPLPO* | Housekeeping | 5’-GGCTGTGGTGCTGATGGGCAAGAA-3’ |
|  |  |  | 3’-TTCCCCCGGATATGAGGCAGCAGT-5’ |
| 3 | *PUM1* | Housekeeping | 5’-CCGGAGATTGCTGGACATATAA-3’ |
|  |  |  | 3’-TGGCACGCTCCAGTTTC-5’ |
| 4 | *ESR1* | Luminal | 5’-CCCACCAGAGGCCCTCGAAA-3’ |
|  |  |  | 3’-AAGCGGGTCACCTGGTCAGT-5’ |
| 5 | *KRT8* | Luminal | 5’-ACAAGGTAGAGCTGGAGTCTCG-3’ |
|  |  |  | 3’-AGCACCACAGATGTGTCCGAGA-5’ |
| 6 | *KRT5* | Basal | 5’-GGAGGTATCCAAGAGGTCACTG-3’ |
|  |  |  | 3’-GGAGGCAAACTTATTGTTGAGG-5’ |
| 7 | *KRT14* | Basal | 5’-CAGTTCTCCTCTGGATCGCA-3’ |
|  |  |  | 3’-CATCGTGCACATCCATGACC-5’ |
| 8 | *GRHL2* | Epithelial | 5’-AGCTACGTCTGGGTGCAGTA-3’ |
|  |  |  | 3’-GCCTTCCCAGAGCAACCTAA-5’ |
| 9 | *CDH1* | Epithelial | 5’-TCCAACGGGAATGCAGTTGA-3’ |
|  |  |  | 3’-TGGGTGAATTCGGGCTTGTT-5’ |
| 10 | *EPCAM* | Epithelial | 5’-GCCAGTGTACTTCAGTTGGTGC-3’ |
|  |  |  | 3’-CCCTTCAGGTTTTGCTCTTCTCC-5’ |
| 11 | *SNAI2* | Mesenchymal | 5’-CCAAACTACAGCGAACTGGA-3’ |
|  |  |  | 3’-GTGGTATGACAGGCATGGAG-5’ |
| 12 | *ZEB1* | Mesenchymal | 5’-GCACCTGAAGAGGACCAGAG-3’ |
|  |  |  | 3’-TGCATCTGGTGTTCCATTTT-5’ |
| 13 | *VIM* | Mesenchymal | 5’-CGTCAGCAATATGAAAGTGTGGCTGC-3’ |
|  |  |  | 3’-CCGGTTGGCAGCCTCAGAGA-5’ |
| 14 | *TWIST1* | Mesenchymal | 5’-CCGGAGACCTAGATGTCATTG-3’ |
|  |  |  | 3’-TTTCCAAGAAAATCTTTGGCATA-5’ |

Table 2: Luminal, epithelial, basal and mesenchymal scores as computed by gene expression analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | LUMINAL SCORE | EPITHELIAL SCORE | BASAL SCORE | MESENCHYMAL SCORE | EPITHELIAL/MESENCHYMAL RATIO |
| MCF7 | 13.2 | 13.0 | 4.9 | 4.3 | 3.1 |
| ZR-75-1 | 11.7 | 13.4 | 6.2 | 5.0 | 3.3 |
| HCC1937 | 10.8 | 12.7 | 11.2 | 8.4 | 1.6 |
| MDA-MB-468 | 11.1 | 12.0 | 8.5 | 7.2 | 1.7 |
| MDA-MB-231 | 6.8 | 5.6 | 4.8 | 11.1 | 0.5 |

**Reference**

1. Nair MG, Desai K, Prabhu JS, Hari PS, Remacle J, Sridhar TS: **β3 integrin promotes chemoresistance to epirubicin in MDA-MB-231 through repression of the pro-apoptotic protein, BAD.** *Experimental Cell Research* 2016, **346:**137-145.

2. Korlimarla A, Prabhu JS, Anupama CE, Remacle J, Wahi K, Sridhar TS: **Separate quality-control measures are necessary for estimation of RNA and methylated DNA from formalin-fixed, paraffin-embedded specimens by quantitative PCR.** *J Mol Diagn* 2014, **16:**253-260.

**Author contributions for the in-vitro experiments:**

Madhumathy G Nair1 , Chandrakala M Naidu1 , Apoorva D Mavatkar1, Jyothi S Prabhu1

1Division of Molecular Medicine, St. John’s Research Institute, St. John’s Medical College, Bangalore, Karnataka, India

MGN was involved in design of the work, the acquisition, analysis, and interpretation of data, and has drafted the work; CMN and ADM have contributed equally towards acquisition, analysis and drafting; JSP was involved in interpretation of data.