

# **PREDICTION OF PROGNOSIS AND CHEMOSENSITIVITY IN BREAST CANCER**

**A DISSERTATION SUBMITTED TO  
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE  
OF BILKENT UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF  
DOCTOR OF PHILOSOPHY  
IN  
MOLECULAR BIOLOGY AND GENETICS**

**BY  
Muhammad Waqas Akbar  
September 2020**

PREDICTION OF PROGNOSIS AND CHEMOSENSITIVITY IN BREAST  
CANCER

By Muhammad Waqas Akbar

September 2020

We certify that we have read this dissertation and that in our opinion it is fully adequate,  
in scope and in quality, as a thesis for the degree of Doctor of Philosophy.



---

Ali Osmay Güre (Advisor)

---

Işık Yuluğ

---

Serkan İsmail Göktuna

---

Burçak Vural

---

Sreeparna Banerjee

Approved for the Graduate School of Engineering and Science:

---

Ezhan Karaşan

Director of the Graduate School

# **ABSTRACT**

## **PREDICTION OF PROGNOSIS AND CHEMOSENSITIVITY IN BREAST CANCER**

**Muhammad Waqas Akbar**

**Ph.D. in Molecular Biology and Genetics**

**Supervisor: Ali Osmay Güre**

**September 2020**

Breast cancer (BC) is responsible for the highest mortality and morbidity out of all the cancers in women which is primarily due to both inter and intra-tumoral molecular heterogeneity. This heterogeneity arises from stemness, epithelial to mesenchymal transition and the type of treatment given to patients. These three biological processes are highly related with each other. Traditional therapy when given to breast cancer patients generally results in the transition of epithelial cells to mesenchymal phenotype. Because treatment targets primarily non-stem cells, it can leave stem cells alive which can later result in a relapse of cancer. In this study we aimed to identify such markers that could classify breast cancer patients into stem/mesenchymal or non-stem/epithelial like phenotypes, to determine how generalized the above stated hypotheses are. We developed a gene list of 15 genes we term as CSC/non-CSC gene list (CNCL) which classifies tumors into stemness and/or EMT based phenotypes and can also classify tumor cells based on their relative sensitivity to treatment with traditional therapeutics such as paclitaxel and doxorubicin. When classified into stem/mesenchymal (CS/M) and non-stem/epithelial (NS/E) phenotypes, we showed that Lapatinib and Midostaurin have a specific growth inhibitory effects on NS/E cells, and CS/M cells, respectively. Surprisingly the CNCL showed prognostic significance only for patients who were treated with paclitaxel in neo-adjuvant setting, while it could not prognosticate most other BC cohorts. We argue that this is due to the dynamic plasticity of these tumors, as studied within the third aim of this thesis.

Secondly, we aimed to identify chemotherapy biomarkers for paclitaxel, cisplatin and doxorubicin to stratify patients in groups that will or will not benefit from these drugs. Using biomarkers, we selected for this purpose, we performed linear regression analysis using breast cancer cell lines to generate cytotoxicity prediction models which can predict IC50 values for these drugs, based on the expression of two genes in each model. Two models were selected for doxorubicin and cisplatin, and three models were selected for paclitaxel. All models were validated both *in silico* and *in vitro*.

Thirdly, we aimed to evaluate breast cancer plasticity that occurs upon treatment or when a tumor metastasizes. We noted that some breast tumors not only switch their clinical subtype but also change their molecular subtype upon treatment or metastasis. As breast cancer patient treatment in the routine practice is routed based on breast cancer subtype, it is very important to identify the subtype switches which can be critical for changes in treatment decisions. Additionally, we also identified metastatic biomarkers using large number of cohorts.

Lastly, as CNCL genes did not show any prognostic importance in terms of both overall survival and metastasis free survival, we checked if the same is true for melanoma. We used Melanin A (MLANA) and Inhibin (INHBA) genes as the markers for invasive/proliferative, stem/non-stem and mesenchymal/epithelial phenotypes. High INHBA expression, which is epithelial, proliferative and non-stem phenotype biomarker, was associated with poor survival and high MLANA expression, which is mesenchymal, invasive and stem phenotype marker, was associated with good prognosis in melanoma patients. Therefore, these findings in melanoma supported our results in breast cancer.

**Keywords:** Breast cancer, Stem cell, Paclitaxel, Cisplatin, Doxorubicin, Biomarker, Metastasis, Plasticity, INHBA, MLANA

# ÖZET

## Meme Kanserinde Prognoz ve Kimyasal Duyarlılık Tahmini

Muhammad Waqas Akbar

Moleküler Biyoloji ve Genetik, Doktora

Tez Danışmanı: Ali Osmay Güre

Eylül 2020

Meme kanseri kadınlarda görülen en yüksek hastalık ve ölüm oranına sahip kanser tipi olup, bu durumun başlıca sebepleri tümör-içi ve tümörler arası moleküler heterojenitedir. Kök hücre özellikleri, epitel-mezenkimal geçiş (EMT) ve hastalara verilen tedaviler bu heterojeniteye sebep olabilen etkenlerdir. Bu üç biyolojik süreç birbirleriyle yüksek oranda ilişkililerdir. Bilinen geleneksel tedavi yöntemleri, hastalara uygulandığında, epitelden mezenkimal fenotipe geçişe neden olmaktadır. Uygulama öncelikle kök hücre olmayan hücreleri hedef aldığından dolayı, kök hücreleri hayatta kalmakta ve daha sonrasında hastalığın nüksetmesine neden olabilmektedirler. Yukarıda verilen hipotezlerin geçerliliklerini görmek adına, bu çalışmada meme kanseri hastalarını kök hücre/mezenkimal ve kök hücre olmayan/epitel fenotipler olarak ayırabilecek belirteçlerin tanımlanmasını amaçladık. Kanser kök hücresi/ kanser kök hücresi olmayan fenotipler şeklinde tanımlanan ve 15 genden oluşan bir gen listesi oluşturulmuştur (CNCL). Bu listenin, tümörleri kök hücre ve/veya EMT özellikli fenotipler olarak sınıflandırabildiği, ve aynı zamanda tümör hücrelerini paklitaksel ve doksorubisin gibi geleneksel ilaç uygulamalarına karşı gösterdikleri göreceli hassasiyetler ölçütünde ayırt edebildiği görülmüştür. Hücre hatları kök hücre/mezenkimal (CS/M) ve kök hücre olmayan/epitel (NS/E) fenotiplere göre sınıflandırıldığında, sırasıyla Lapatinib'in NS/E hücreleri üzerinde ve Midostaurin'in CS/M hücreleri üzerinde spesifik büyüme inhibisyonu sağladıklarını bulgadık. Şaşırtıcı bir şekilde, CNCL'in prognoz ile istatistiksel olarak anlamlı ilişkisi sadece Paklitaksel temelli neo-adjuvant tedavisi alan hastalar için geçerliken, geriye kalan birçok meme kanseri kohortu CNCL ile prognoz edilememiştir. Bu tezin üçüncü amacında

da değinildiği üzere, bu durumun ilgili tümörlerin dinamik plastisitelerinden kaynaklandığına inanıyoruz.

İkincil olarak, paklitaksel, sisplatin ve doksorubisin ilaçları ile tedavi gören hastaların tedavi yanıtlarının tahmin edilmesine yarayabilecek kemoterapi biyobelirteçlerinin tespitleri hedeflenmiştir. Meme kanseri hücre hatlarında bu doğrultuda seçtiğimiz belirteçler kullanılarak, yapılan doğrusal regresyon (linear regression) analizi ile ilaçların IC50 değerlerini 2 genin ifadesine bağlı olarak tahmin edebilen bir sitotoksosite tahmin modeli oluşturulmuştur. doksorubisin ve cisplatin için 2 model, paklitaksel için ise 3 model seçilmiştir. Modellerin tamamı *in siliko* ve *in vitro* olarak doğrulanmıştır.

Üçüncü olarak meme kanseri plastisitesinin tedavi sonrası ve tümör metastazı gerçekleşmesi durumlarında değerlendirilmesi amaçlanmıştır. Tedavi ve metastaz sonrası bazı meme tümörlerinin klinik alt türlerini değiştirmelerinin yanı sıra moleküler alt türlerini de değiştirebildikleri gösterilmiştir. Rutin pratikte meme kanseri tedavisi meme kanseri alt türlerine göre yönlendirildiğinden, tedavi kararlarındaki değişiklikler için kritik öneme sahip olabilecek alt tür geçişlerinin tespiti bir hayli önem arz etmektedir. Ek olarak, yüksek sayıda hasta kohortu kullanılarak metastatik biyobelirteçler de tanımlanmıştır.

Son olarak, çalışmamızda CNCL gen listesinin genel sağkalım ve metastazsız sağkalım açısından prognozistik bir önemi olmadığı görüldüğünden, benzer bir çalışma melanom için tekrarlanmıştır. Melanin A (MLANA) ve İnhibin (INHBA) genleri invazif/proliferatif, kök hücre/kök hücre olmayan ve mezenkimal/epitel fenotipler için belirteç olarak kullanılmıştır. Melanom hastalarında epitel, proliferatif ve kök hücre olmayan fenotip biyobelirteci INHBA'nin yüksek ifadesi kötü prognoz ile ilişkili iken, mezenkimal, invazif ve kök hücre fenotipi belirteci olan MLANA'nın yüksek ifadesi iyi prognoz ile ilişki göstermiştir. Bu sebepten dolayı, melanomdaki bulgularımız meme kanserinde elde ettiğimiz sonuçları desteklemektedir.

**Anahtar sözcükler:** Meme kanseri, Kök hücre, Paklitaksel, Sisplatin, Doksorubisin, Biyobelirteç, Metastaz, Plastisite, INHBA, MLANA

## Acknowledgements

All praises and thanks to Almighty Allah; the ultimate source of all knowledge to mankind and for His endless blessings for humanity

It gives me extreme pleasure to express my deepest gratitude to my indulgent mentor Assoc. Prof. Dr. Ali Osmay Güre without whose guidance and support, I couldn't overcome all the challenges to pursue this dream. He is always generous to share his philosophy, creative ideas and reasoning with us both in and out of the lab. I am very grateful to him for his kind valuable guidance and ever willing support throughout my stay in Turkey. He always inspired me to explore critical questions and conduct quality research through thick and thin. It is an honor and privilege to be a part of his inventive team.

Secondly, I owe my profound regards to Assoc. Prof. Dr. Özgür Şahin, Assoc. Prof. Ozlen Konu, Assoc. Prof. Dr. Işık Yuluğ, Assist. Prof. Serkan Ismail Göktuna, Prof. Dr. Burçak Vural and Prof. Sreeparna Banerjee for their valuable counsel and insights for the preparation of this thesis.

I would like to extend heartiest thanks to all current and previous members of AOG Lab as each one of them is very special to me. Murat İşbilen's sincerity, Can and Seyhan Turk's compassion, Kerem Mert Şenses' guidance, Seçil Demirkol Canli's kindness, Barış Küçükkaraduman's benevolence, Abbas Güven Akçay's perceptiveness, Alper Poyraz's resourcefulness, Şükrü Atakan's inquisitiveness and Ege Dedeoğlu's liveliness was the dearest to me. I am also grateful to our relatively new group members Shila Azizolli, Marzana Ishraq, Isli Cela, Farid Ahadli and Farid Sadati for their help and support.

I would like to express my profound appreciation for the trustworthiness, good will and loyalty shown to me by Adem Demir. I am sincerely thankful to OŞ lab especially Unal Tokat, Pelin Ersan, Oğuzhan Tarman and Nevin Belder. I am also thankful to Murat Yaman and Seniye Targen for their help and sympathy.

Among my Pakistani peers, I am very grateful to Mr. and Mrs. Naveed Mehmood, Mr. and Mrs. Talha Masood and Muhammad Zakwan.

I gratefully acknowledge the funding received from the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 117S058 and Higher Education Commission, Pakistan.

Last but not least, I admit that I owe all my achievements to my sincere and most loving parents and my sisters, whose prayers and motivational support always remained a source of determination for me.

Muhammad Waqas Akbar



**TO MY FAMILY**

## Table of Contents

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| 1. Introduction .....                                                                  | 1  |
| 1.1. Breast cancer and its classification .....                                        | 1  |
| 1.2. Breast cancer stem cells.....                                                     | 1  |
| 1.2.1. Breast CSCs markers .....                                                       | 2  |
| 1.2.2. Cancer stem cells and epithelial to mesenchymal transition.....                 | 2  |
| 1.2.3. Role of biological pathways and associated factors in breast cancer stemness .. | 3  |
| 1.2.4. CSCs resistance to therapy .....                                                | 4  |
| 1.2.5. Presence of CSC in breast cancer cell lines .....                               | 5  |
| 1.2.6. CSCs enrichment in vitro .....                                                  | 5  |
| 1.3. Chemotherapy in BC .....                                                          | 6  |
| 1.3.1. Paclitaxel .....                                                                | 7  |
| 1.3.2. Doxorubicin.....                                                                | 8  |
| 1.3.3. Cisplatin .....                                                                 | 9  |
| 1.4. Breast cancer plasticity .....                                                    | 10 |
| 1.5. Metastatic biomarkers for BC .....                                                | 12 |
| 1.6. Role of MLANA and INHBA in Melanoma .....                                         | 15 |
| 2. Aim of the study .....                                                              | 17 |
| 3. Materials.....                                                                      | 19 |
| 3.1. General laboratory reagents .....                                                 | 19 |
| 3.2. General laboratory equipment.....                                                 | 19 |
| 3.3. Cell culture reagents.....                                                        | 20 |
| 3.4. Cell culture media and solutions.....                                             | 20 |
| 4. Methods.....                                                                        | 22 |
| 4.1. Cell culture conditions.....                                                      | 22 |
| 4.2. Culturing cryopreserved cells .....                                               | 22 |
| 4.3. Cell culture.....                                                                 | 22 |
| 4.4. Cell cryopreservation .....                                                       | 22 |
| 4.5. Mammosphere culture (3D cell culture).....                                        | 23 |
| 4.6. RNA extraction .....                                                              | 23 |
| 4.7. DNase treatment.....                                                              | 24 |
| 4.8. RNA quantification .....                                                          | 24 |

|                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.9. cDNA synthesis.....                                                                                                           | 24 |
| 4.10. End point polymerase chain reaction .....                                                                                    | 25 |
| 4.11. Agarose gel electrophoresis .....                                                                                            | 25 |
| 4.12. Quantitative Real time Polymerase Chain Reaction (qRT-PCR) .....                                                             | 26 |
| 4.13. Drug cytotoxicity .....                                                                                                      | 26 |
| 4.14. siRNA Knockdown and Cell viability.....                                                                                      | 26 |
| 4.15. <i>In silico</i> analyses .....                                                                                              | 27 |
| 4.16. Statistical analysis .....                                                                                                   | 27 |
| 5. Results.....                                                                                                                    | 29 |
| 5.1. Classification of breast cancer based on stemness and EMT .....                                                               | 29 |
| 5.1.1. Robustness of the CSC/non-CSC gene list (CNCL) classification .....                                                         | 29 |
| 5.1.2. Altered Molecular Mechanisms as indicated by differentially expressed genes<br>between CSC and non-CSC like cell lines..... | 32 |
| 5.1.3. CNCL can identify stemness <i>in silico</i> .....                                                                           | 35 |
| 5.1.4. CNCL can identify EMT <i>in silico</i> .....                                                                                | 41 |
| 5.1.5. CNCL can predict drug sensitivity <i>in silico</i> .....                                                                    | 44 |
| 5.1.6. Stemness score (SS) based quantitative analyses .....                                                                       | 47 |
| 5.1.7. CNCL validation <i>in vitro</i> .....                                                                                       | 51 |
| 5.1.8. Prognostic significance of CNCL .....                                                                                       | 56 |
| 5.2. Chemosensitivity markers of paclitaxel, cisplatin and doxorubicin in breast cancer<br>.....                                   | 59 |
| 5.2.1. Paclitaxel biomarkers.....                                                                                                  | 61 |
| 5.2.2. Cisplatin biomarkers .....                                                                                                  | 65 |
| 5.2.3. Doxorubicin biomarkers.....                                                                                                 | 67 |
| 5.3. Breast cancer plasticity .....                                                                                                | 71 |
| 5.3.1. PAM50 plasticity .....                                                                                                      | 71 |
| 5.4. Metastatic biomarkers .....                                                                                                   | 76 |
| 5.4.1. Prognostic importance of hematopoietic cells for metastasis .....                                                           | 79 |
| 5.5. Prognostic importance of INHBA and MLANA in melanoma .....                                                                    | 82 |
| 6. Discussion .....                                                                                                                | 94 |
| 6.1. CNCL based classification of breast cancer into stemness and EMT subtypes .....                                               | 94 |
| 6.2. Breast cancer plasticity.....                                                                                                 | 96 |

|                                                                                               |     |
|-----------------------------------------------------------------------------------------------|-----|
| 6.3. Chemosensitivity markers of paclitaxel, cisplatin and doxorubicin in breast cancer ..... | 98  |
| 6.3.1. Paclitaxel biomarkers.....                                                             | 99  |
| 6.3.2. Cisplatin Biomarkers .....                                                             | 100 |
| 6.3.3. Doxorubicin Biomarkers.....                                                            | 101 |
| 6.4. Biomarkers for metastasis.....                                                           | 102 |
| 6.4.1. Good prognostic genes.....                                                             | 102 |
| 6.4.2. Bad prognostic genes .....                                                             | 106 |
| 6.4.3. Prognostic importance of hematopoietic cells for metastasis .....                      | 108 |
| 6.5. Prognostic importance of INHBA and MLANA in melanoma .....                               | 110 |
| 7. Future Perspectives.....                                                                   | 112 |
| 7.1. Classification of Breast cancer based upon stemness and EMT.....                         | 112 |
| 7.2. Chemosensitivity markers of paclitaxel, doxorubicin and cisplatin in breast cancer ..... | 113 |
| 7.3. Breast cancer plasticity.....                                                            | 113 |
| 7.4. Biomarkers for metastasis and the role of hematopoietic cells in metastasis.....         | 114 |
| 7.5. Role of INHBA and MLANA in melanoma.....                                                 | 115 |
| 8. References .....                                                                           | 116 |
| 9. Appendix.....                                                                              | 137 |
| 10. Publications .....                                                                        | 144 |

## List of Figures

|                                                                                                                                                                                          |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 5.1: E cadherin (CDH1) expression difference between CSC and non-CSC like clusters. ....                                                                                          | 35 |
| Figure 5.2: CNCL clustering in dataset GSE36643. ....                                                                                                                                    | 39 |
| Figure 5.3. Clustering of primary breast cancer cells sorted by ALDH1 expression with CNCL.....                                                                                          | 41 |
| Figure 5.4: CNCL clustering in dataset GSE9691. ....                                                                                                                                     | 42 |
| Figure 5.5: CNCL clustering in dataset GSE24202. ....                                                                                                                                    | 43 |
| Figure 5.6: CNCL clustering in dataset GSE7515. ....                                                                                                                                     | 44 |
| Figure 5.7: CNCL clustering in GSE24460.....                                                                                                                                             | 45 |
| Figure 5.8: CNCL clustering in GSE10281. CNCL clustered the majority of pre letrozole treated samples as NS/E and post treated samples as CS/M.....                                      | 46 |
| Figure 5.9: Stemness score-based clustering in dataset GSE12791.....                                                                                                                     | 49 |
| Figure 5.10: Stemness score-based clustering in dataset GSE23399.....                                                                                                                    | 49 |
| Figure 5.11: Stemness score-based clustering in dataset GSE16179.....                                                                                                                    | 50 |
| Figure 5.12: Delta(r) based analysis in GSE28844. ....                                                                                                                                   | 51 |
| Figure 5.13: Midostaurin shows preferential cytotoxicity for CS/M cell lines and Lapatinib shows preferential cytotoxicity to NS/E cell lines.....                                       | 53 |
| Figure 5.14: Spheroid cytotoxicity alteration due to phenotypic switch. MDA-MB-157 showed increased sensitivity to Midostaurin and resistance to Lapatinib upon spheroid generation..... | 54 |
| Figure 5.15: SiZEB1 and SiSNAI2 knockdown efficiency in MDA-MB-157 cell line....                                                                                                         | 55 |
| Figure 5.16: ZEB1 and SNAI2 knockdown resulted in increased sensitivity to Lapatinib and resistance to Midostaurin in MDA-MB-157 cell line as expected. ....                             | 55 |
| Figure 5.17: Survival analysis of GSE25066.....                                                                                                                                          | 58 |
| Figure 5.18: Delta (r) based distribution of breast cancer patients in molecular subtypes in GSE25066. ....                                                                              | 58 |
| Figure 5.19: Paclitaxel, cisplatin and doxorubicin cytotoxicity data <i>in vitro</i> . ....                                                                                              | 60 |
| Figure 5.20: Paclitaxel, cisplatin and doxorubicin cytotoxicity data <i>in silico</i> . ....                                                                                             | 61 |
| Figure 5.21: Pearson correlation analysis between paclitaxel predicted IC50 and <i>in vitro</i> cytotoxicity data for selected 3 models. ....                                            | 62 |
| Figure 5.22: Validation of 3 paclitaxel models in dataset GSE12791.....                                                                                                                  | 63 |
| Figure 5.23: Validation of 3 paclitaxel models in dataset GSE23399.....                                                                                                                  | 64 |
| Figure 5.24: Validation of 3 paclitaxel models in dataset GSE28784.....                                                                                                                  | 64 |
| Figure 5.25: Validation of 2 cisplatin models with <i>in vitro</i> cisplatin cytotoxicity data. ...                                                                                      | 66 |
| Figure 5.26: Validation of 2 cisplatin models in dataset GSE28274. ....                                                                                                                  | 66 |
| Figure 5.27: Validation of 2 doxorubicin models with <i>in vitro</i> cytotoxicity data.....                                                                                              | 68 |
| Figure 5.28: Validation of 2 doxorubicin models in dataset GSE13477.....                                                                                                                 | 68 |
| Figure 5.29: Validation of 2 doxorubicin models in dataset GSE24460.....                                                                                                                 | 69 |
| Figure 5.30: Validation of 2 doxorubicin models in dataset GSE72362.....                                                                                                                 | 70 |
| Figure 5.31: Validation of 2 doxorubicin models in dataset GSE3929.....                                                                                                                  | 70 |
| Figure 5.32: Biomarkers for breast cancer metastasis .....                                                                                                                               | 79 |

|                                                                                                                   |    |
|-------------------------------------------------------------------------------------------------------------------|----|
| Figure 5.33: Prognostic evaluation of INHBA in TCGA cohort. ....                                                  | 84 |
| Figure 5.34: Prognostic evaluation of MLANA in TCGA cohort. ....                                                  | 84 |
| Figure 5.35: Prognostic evaluation of INHBA in GSE65904 cohort for distant metastasis<br>free survival data. .... | 87 |
| Figure 5.36: Prognostic evaluation of MLANA in GSE65904 cohort for distant metastasis<br>free survival data. .... | 87 |
| Figure 5.37: Prognostic evaluation of INHBA in GSE65904 cohort for disease specific<br>survival data. ....        | 89 |
| Figure 5.38: Prognostic evaluation of MLANA in GSE65904 cohort for disease specific<br>survival data. ....        | 89 |
| Figure 5.39: Prognostic evaluation of MLANA in GSE59455 cohort for overall survival<br>data.....                  | 91 |
| Figure 5.40: Prognostic evaluation of MLANA in GSE54467 cohort for overall survival<br>data.....                  | 93 |

## List of Tables

|                                                                                                                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 3.1: General laboratory reagents and vendors .....                                                                                                                                                           | 19 |
| Table 3.2: Cell culture media and reagents.....                                                                                                                                                                    | 20 |
| Table 5.1: CSC/non-CSC gene list (CNCL) .....                                                                                                                                                                      | 29 |
| Table 5.2: CNCL classification robustness across different datasets and platforms. N/A shows that the cell line was not present in that dataset.....                                                               | 30 |
| Table 5.3: Gene sets upregulated in CSC like (CS/M) cell lines .....                                                                                                                                               | 33 |
| Table 5.4: Gene sets upregulated in Non-CSC like (NS/E) cell lines. ....                                                                                                                                           | 34 |
| Table 5.5: Differential expression of CD44, CD24 and ALDH in CS/M and NS/E groups .....                                                                                                                            | 36 |
| Table 5.6: CNCL genes showed a significant correlation pattern with CD44, CD24 and ALDH in CCLE and CGP datasets. ....                                                                                             | 36 |
| Table 5.7: CD44 and CD24 correlation with ALDH in CCLE and CGP datasets.....                                                                                                                                       | 37 |
| Table 5.8: CNCL genes correlation with CD44, CD24 and ALDH1.....                                                                                                                                                   | 38 |
| Table 5.9: Pearson correlation analysis between CNCL genes and stemness genes (CD44 and CD24).....                                                                                                                 | 40 |
| Table 5.10: Phenotype comparison of matched tumor samples pre and post treatment with letrozole.....                                                                                                               | 47 |
| Table 5.11: CNCL stemness reference matrix .....                                                                                                                                                                   | 48 |
| Table 5.12: qPCR data was highly concordant with CCLE gene expression data.....                                                                                                                                    | 52 |
| Table 5.13: CNCL genes showed a strong intergenic correlation pattern as expected. ....                                                                                                                            | 52 |
| Table 5.14: Prognostic evaluation of 16 breast cancer cohorts based upon SS. ....                                                                                                                                  | 57 |
| Table 5.15: Multivariate analysis of SS delta (r) with clinical parameters such as T stage, N stage and breast cancer molecular subtypes, showed SS delta (r) as a significant independent prognostic marker. .... | 59 |
| Table 5.16: Linear regression cytotoxicity models of paclitaxel. ....                                                                                                                                              | 62 |
| Table 5.17: Linear regression cytotoxicity models of cisplatin. ....                                                                                                                                               | 65 |
| Table 5.18: Linear regression cytotoxicity models of doxorubicin.....                                                                                                                                              | 67 |
| Table 5.19: Comparison of PAM50 subtyping for samples collected through fine needle aspiration and core needle biopsy (GSE32518). ....                                                                             | 72 |
| Table 5.20: Comparison of PAM50 subtypes pre- and post-Letrozole therapy (GSE10281).....                                                                                                                           | 73 |
| Table 5.21: Comparison of PAM50 subtypes pre- and post-Anthracyclines and Taxanes based neoadjuvant chemotherapy (GSE28844). ....                                                                                  | 73 |
| Table 5.22: Comparison of PAM50 subtypes pre- and post- neoadjuvant chemotherapy (GSE32072).....                                                                                                                   | 74 |
| Table 5.23: Comparison of PAM50 subtyping for primary tissues vs. brain metastatic tissues (GSE125989). ....                                                                                                       | 74 |
| Table 5.24: Comparison of PAM50 subtyping for primary tissues vs. metastatic tissues (GSE110590).....                                                                                                              | 76 |
| Table 5.25: Patient number and treatment protocols for cohorts used for metastasis biomarker analysis.....                                                                                                         | 77 |

|                                                                                                                                                                                                                                    |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Table 5.26: Hazard ratios (HR) of hematopoietic cells for metastasis free survival. Cells highlighted in green show good prognosis and red shows bad prognosis. Cells with thick borders show HR with significant Cox p value..... | 80        |
| Table 5.27: Cox regression p values for Hazard ratios in Table 5.26. ....                                                                                                                                                          | 81        |
| Table 5.28: Cox regression analysis for INHBA, MLANA and clinical parameters.....                                                                                                                                                  | 83        |
| Table 5.29: Patient distribution, median survival and Log rank p value for INHBA and MLANA classifications. ....                                                                                                                   | 84        |
| Table 5.30: Multivariate analysis for TCGA with parameters significant with univariate analysis.....                                                                                                                               | 85        |
| Table 5.31: Cox regression analysis for INHBA and MLANA for GSE65904 DMFS ....                                                                                                                                                     | 86        |
| Table 5.32: Patient distribution, median survival and Log rank p value for INHBA and MLANA classifications in GSE65294 DMFS.....                                                                                                   | 88        |
| Table 5.33: Cox regression analysis for INHBA and MLANA for GSE65904 DSS .....                                                                                                                                                     | 88        |
| Table 5.34: Patient distribution, median survival and Log rank p value for INHBA and MLANA classifications in GSE65904 DSS. ....                                                                                                   | 90        |
| Table 5.35: Cox regression analysis for INHBA, MLANA, BRAF mutation, NRAS mutation, and Berslow depth in dataset GSE59455. ....                                                                                                    | 90        |
| Table 5.36: Patient distribution, median survival and Log rank p value for MLANA classifications in GSE59455 OS.....                                                                                                               | 91        |
| <b>Table 5.37: Multivariate analysis for MLANA classification and Berslow depth in dataset GSE59455.....</b>                                                                                                                       | <b>92</b> |
| <b>Table 5.38: Cox regression analysis for MLANA in GSE54467 .....</b>                                                                                                                                                             | <b>92</b> |
| <b>Table 5.39: Patient distribution, median survival and Log rank p value for MLANA classifications in GSE54467 OS.....</b>                                                                                                        | <b>93</b> |
| Table 9.1: Datasets used for validation of CNCL gene list. ....                                                                                                                                                                    | 137       |
| Table 9.2: Datasets used in prognostic evaluation for CNCL gene list.....                                                                                                                                                          | 141       |

## Abbreviations

|      |                                           |
|------|-------------------------------------------|
| BC   | Breast Cancer                             |
| CSC  | Cancer Stem Cells                         |
| CNSC | Cancer Non-Stem Cells                     |
| FACS | Fluorescence-activated Cell Sorting       |
| EMT  | Epithelial to Mesenchymal Transition      |
| MET  | Mesenchymal to Epithelial Transition      |
| LumA | Luminal A                                 |
| LumB | Luminal B                                 |
| CNCL | CSC/non-CSC gene list                     |
| CS/M | Cancer Stem/Mesenchymal Phenotype         |
| NS/E | Non Stem/Epithelial Phenotype             |
| NA   | Not available                             |
| TNBC | Triple Negative Breast Cancer             |
| IC50 | The half maximal Inhibitory concentration |

## **1. Introduction**

### **1.1. Breast cancer and its classification**

Breast cancer (BC) is one of the cancers responsible for the highest mortality and morbidity in women. According to cancer statistics 2019 in the USA only, new estimated cases for breast cancer were the highest among all cancer patients (271270 cases). Similarly, for BC, mortality was estimated to be the second highest after lung cancer in women (41760 deaths in women) (Siegel et al., 2019). BC is a heterogeneous disease and is clinically classified based upon immunohistochemistry expression of estrogen receptor (ER), Progesterone receptor (PR) and human epidermal growth factor receptor (HER2, also known as ERBB2) expression. For both ER and PR, nuclear staining positive for >1% tumor cells, is considered positive by the American Society of Clinical Oncology (ASCO) but some countries such as France recommend 10% as the cutoff for these receptors. Additionally, by recommendations of ASCO, HER2 status is confirmed via two techniques at least and for this purpose approaches like immunohistochemistry (IHC), fluorescent *in situ* Hybridization and chromogenic *in situ* hybridization is used. Immunohistochemistry is the most preferred choice worldwide and 10% staining is used as the cutoff for HER2 staining (Harbeck et al., 2019). Moreover, BC is also molecularly classified into Luminal A (LumA), Luminal B (LumB), HER2 enriched (HER2), Basal and Normal like types. This molecular classification is being widely used and is known as PAM50 (van 't Veer et al., 2002; Parker et al., 2009). Apart from these classifications, BC can also be divided into cancer stem cells (CSC) and cancer non-stem cells (CNSC) which further shows the extremely heterogeneous nature of breast cancer (Hii et al., 2020).

### **1.2. Breast cancer stem cells**

BC is thought to develop through the random accumulation of the right combinations of mutations in breast cancer cells. However, BC research has extensively supported the cancer stem cell hypothesis indicating two different kinds of populations which are CSC and CNSC to be present in tumors (Kakarla et al., 2012). There are two proposed models through which breast cancer stem cells are generated. The first model is hierarchical cancer stem cell model which states that CSCs are generated through cancerous transformation of

normal breast stem cells or early progenitor cells, and after transformation, these cells differentiate into CNSCs. Additionally, CNSCs are also generated through transformation of normal differentiated breast cells. The second model states that NCSCs generate CSCs when undergoing de-differentiation (Butti et al., 2019). Breast CSCs were first reported by Al hajj et al. in 2003. The authors identified cells that were CD44<sup>+</sup>/CD24<sup>-</sup>. These cells were sorted through fluorescence-activated cell sorting (FACS) and when as low as 200 cells were injected in NOD/SCID mice, highly aggressive tumors were generated. While injections of other cells which were CD24<sup>+</sup>, did not develop any tumors even at high numbers such as 20000 cells (Al-Hajj et al., 2003). The phenomenon of CSCs self-renewal along with the capability to differentiate, through which they generate phenotypically heterogeneous CNSCs, is known as asymmetrical division. Additionally, CSCs are considered highly important due to their significant role in malignancy, metastasis and relapse of BC (Velasco-Velázquez et al., 2012).

#### **1.2.1. Breast CSCs markers**

In breast cancer research, due to the heterogeneous nature of BC, many different markers have been used to identify and sort CSCs from CNSCs. The most accepted markers among these are CD44<sup>+</sup>/CD24<sup>-</sup> which were first described by Al-hajj et al (Al-Hajj et al., 2003). Others markers used for CSC identification include ALDH1, CD90, CD38, ABCB5, CD34, CD49f, GD2 (Chen et al., 2013a), CD133, EpCAM, PKH, CD271 and ANTXR1 (Battula et al., 2012; Vieira et al., 2012; Liu et al., 2014a; Islam et al., 2015; Zhao et al., 2017). These markers are used either individually or in combination with each other in breast cancer stem biology studies. The most important technique to sort CSCs from CNSCs is FACS. Other techniques used to characterize CSCs include RT-qPCR, IHC, and Magnetic-Activated Cell Sorting (MACS)(Islam et al., 2015).

#### **1.2.2. Cancer stem cells and epithelial to mesenchymal transition**

Epithelial to mesenchymal transition (EMT), as the name indicates, is the process through which cells with epithelial properties are transformed into mesenchymal cells. This phenomenon is of high importance as it plays a crucial part in embryogenesis and developmental biology. Additionally, the phenomenon where mesenchymal cells

transform into epithelial cells is termed as mesenchymal to epithelial transition (MET). Both EMT and MET play an important role in breast cancer development and metastasis. These phenotypes naturally co-exist as well in breast cancer tumors (Takebe et al., 2011). Importantly it has been reported that both CSCs and mesenchymal cancer cells generated through EMT show similar characteristics. These characteristics include the capability of tumor generation from minimal cells, resistance to chemotherapy and radiotherapy among other treatments, immunity to apoptosis, metastasis generation, migration and loss of differentiation (Floor et al., 2011). Interestingly ectopic expression of mesenchymal markers such as TWIST and SNAIL generates cells that also express stem cell markers such as CD44<sup>+</sup>/CD24<sup>-</sup>. Moreover, these cells upon proliferation generate both CSCs and CNSCs (Mani et al., 2008). Similarly, breast tumors with mesenchymal phenotype express stem cell markers as well both in mouse and human patients (Puisieux et al., 2014). Another important common mechanism expressed by both stem and mesenchymal cells is resistance to anoikis. Epithelial cells rely on interactions between cell and matrix for their survival and if these cells are grown in suspension where no anchorage is available, this results in programmed cell death termed as anoikis. To survive, these cells undergo EMT and develop a mesenchymal phenotype that is resistant to anoikis. Interestingly similar to mesenchymal cells, stem cells are also resistant to anoikis. Additionally, resistance to anoikis is necessary for metastasis (Frisch et al., 2013). These shared similarities between mesenchymal cancer cells and stem cells lead to conclusion that both processes might be induced and regulated by related biological pathways.

### **1.2.3. Role of biological pathways and associated factors in breast cancer stemness**

Various signaling pathways such as Notch, Wnt, Hedgehog (Hh) and Hippo pathways show alterations in breast CSCs and manage the survival and maintenance of these cells (Clara et al., 2019). Wnt signaling plays an important role in the development of breast tissue and its deregulation is critical for breast cancer. In basal breast cancer, Wnt signaling is associated with worse prognosis and increased Wnt activity is associated with the increased tumor-initiating capacity of CSCs (which is a hallmark of CSCs) (Chakrabarti et al., 2014). Ras/MAPK signaling is another important mechanism whose activation in CNSCs results in the transition of those cells in CSC phenotype (Morel et al., 2008). Stem

cell renewal pathways such as Notch signaling play an important part in stemness and breast stem cells cultured in mammosphere culture show increased Notch receptors (Farnie and Clarke, 2007). Hh signaling is responsible for cellular polarity and plays an important part in stem cell maintenance (Takebe et al., 2011). Additionally, CSCs show increased expression of HH pathway components such as PTCH1, GLI1 and GLI2 when compared with CNSCs (Kasper et al., 2009). Hippo pathway is also an important regulator of stemness as TAZ, a transducer of Hippo pathway, is essential for sustaining tumor-initiating properties of CSCs and also plays an important role in metastatic potential and chemoresistance of these cells (Cordenonsi et al., 2011; Bartucci et al., 2015).

Several important factors have previously been reported for their importance in the induction of stemness in BC. In several studies, authors treated CNSCs with factors leading to their transition in CSCs.  $\text{TNF}\alpha$  and  $\text{TGF}\beta$  treatment over prolonged time converted CNSC into CSCs through EMT. Similarly, IL6 and IL8 treatment has been previously shown to induce stemness and chemoresistance in breast cancer cells (Chen et al., 2018). Moreover in one study, the serum levels of IL8 positively correlated with metastasis in BC. In another study, CCL2 (another cytokine) was shown to induce the self-renewing capacity of CSCs but this was not accompanied by CSCs differentiation to CNSCs. And inhibition of CCL2 in CSCs significantly inhibited tumorigenesis and NOTCH1 expression (Chin and Wang, 2014).

#### **1.2.4. CSCs resistance to therapy**

Breast cancer stem cells are highly resistant to radiotherapy and are enriched after radiotherapy (Phillips et al., 2006). Lagadec et al. showed that non-tumorigenic cells upon treatment with ionizing radiation, developed a CSC phenotype with increased expression of OCT4 and SOX2 (Lagadec et al., 2012). Several studies have shown that CSCs are also resistant to traditional chemotherapies. When breast cancer is treated with these therapies, mostly CNSC cells are targeted and CSCs escape this treatment and over time, these cells ultimately lead to the relapse of tumors. Gupta et al. showed that when breast cancer cells are treated with paclitaxel over a prolonged period of time, this results in increased CSC population in surviving cells (Gupta et al., 2009b). Van Phuc et al. showed that CSC cells

are resistant to doxorubicin and upon downregulation of CD44, these cells become sensitive to chemotherapeutics (Van Phuc et al., 2011). Similar results have been shown for tamoxifen, fulvestrant, cisplatin, 5-fluorouracil and gemcitabine (Fillmore and Kuperwasser, 2008; Shafee et al., 2008; Wang et al., 2012; Bhat-Nakshatri et al., 2013; Samanta et al., 2014). It is important to target both CSC and CNSC populations using combination therapy so that both populations can be eliminated from tumors (Liu and Wicha, 2010). In this regard Gupta et al. have previously shown that paclitaxel is effective against CNSC and salinomycin is effective against CSC population in breast cancer cells (Gupta et al., 2009b). Additionally, other compounds that can target Notch, Wnt, Hedgehog (Hh) and Hippo pathways can be identified and used in combination with traditional therapies to eliminate breast cancer.

#### **1.2.5. Presence of CSC in breast cancer cell lines**

CSCs are not only present in breast tumors but established breast cancer cell lines show the presence of these cells. The presence of CSCs in breast cancer cell lines varies among molecular classes. Cell lines can be divided into 3 groups based upon the presence of CD44<sup>+</sup>/CD24<sup>-</sup> stem cell marker, first group resembles the luminal subtype and shows the least stem cells (e.g. MCF7), and rest two groups are basal where the first one shows basal-like morphology with reduced expression of stem cells (e.g. SUM149) while the other one has basal morphology and more than 90% of its cells are stem cells (e.g. MDA-MB-231) (Fillmore and Kuperwasser, 2008). Similarly using ALDH1 marker, CSCs can be identified in breast cancer cell lines and xenografts as well (Charafe-Jauffret et al., 2009). These studies suggest that breast cancer cell lines can be a suitable model for studying CSCs.

#### **1.2.6. CSCs enrichment in vitro**

Cancer cells who lose contact with surrounding cells or extracellular surface, undergo apoptosis known as anoikis. Resistance to this biological trait is important for the cells to achieve metastatic potential (Simpson et al., 2008). *In vitro*, this trait is exploited to enrich cancer cells for stemness. CSCs cells are enriched in mammosphere culture (3D cell culture). CNSCs require a support surface for their attachment and cannot survive

mammosphere culture as low attachment flasks are usually used for this purpose. Only CSCs can survive this culture because of their resistance to anoikis and growth in spheres. Importantly if mammospheres are cultured in normal flasks as monolayers, they give rise to both CSC and CNSC populations (Grimshaw et al., 2008; Shaw et al., 2012; Wang et al., 2014). Alternative methods to mammosphere culture in low attachment flasks are hanging drop method, rotary culture and culture on matrigel (Vinci et al., 2012). Mammosphere is a cheaper method when compared with other methods but its biggest caveat is the requirement of several growth factors and a prolonged time for establishing culture. CSCs are also grown *in vivo* as a small number of cells are transplanted in mice which give rise to tumors resembling original tumor tissue. Limitations of this method include a lack of cytokines and altered niche (Valent et al., 2012).

### **1.3. Chemotherapy in BC**

BC tumors are treated with chemotherapy in two settings, neoadjuvant therapy where patients are treated with chemotherapy to decrease the size of tumor for probable surgery, and adjuvant setting where patients are treated with chemotherapy after the surgery. Early breast cancer Luminal A and B subtype tumors are treated with endocrine therapy after surgery and/or radiotherapy if the risk of recurrence is low, but if this risk is high then these patients are treated with a combination of either anthracycline and taxane or docetaxel and cyclophosphamide. In HER2 subtype of early breast cancer, patients are treated with chemotherapy (adjuvant paclitaxel and trastuzumab for 1 year) if the pathological (P) stage is 1 but if it is above 1 then those patients are treated with dual anti-HER2 therapy (trastuzumab and pertuzumab) along with a combination of either anthracycline and paclitaxel or docetaxel, and cisplatin in neoadjuvant setting. For these patients, if pCR is achieved after surgery then they are treated with anti-HER2 therapy for 1 year and if pCR is not achieved then these are treated with T-DM1. For early triple-negative breast cancer (TNBC) or Basal subtype, chemotherapy is preferred in the neoadjuvant setting with anthracycline and paclitaxel (platinum-based chemotherapy might also be added) and if after surgery pCR is not achieved then those patients are treated with capecitabine. For advanced breast cancer, Luminal A and B subtype are treated with adjuvant endocrine therapy and chemotherapy options include anthracycline, paclitaxel, vinorelbine and

eribulin. For HER2 subtype, first-line adjuvant treatment includes trastuzumab and pertuzumab along with chemotherapy (taxane or vinorelbine), and the second line of treatment is T-DM1. Advanced TNBC or Basal subtype is treated with nab-paclitaxel plus atezolizumab. For some patients, platinum-based therapy is also a good option (Harbeck et al., 2019). Hence paclitaxel (taxane family), cisplatin (platinum-based drug) and doxorubicin (anthracycline family) are most commonly used in breast cancer treatment, so it's important to identify therapeutic biomarkers for these drugs.

### **1.3.1. Paclitaxel**

Paclitaxel is an antimitotic drug being used in multiple cancer treatments such as lung cancer, prostate cancer, gastric cancer, bladder cancer as well as breast cancer. It binds to microtubules and results in their stabilization by promoting  $\alpha$  and  $\beta$  subunits assembly as these are building blocks of microtubules. This stabilization of microtubule further hampers the cell's ability to divide, halting the cell division in G2 or M phase of cell cycle leading to cell death. This drug was approved for breast cancer treatment in 1994 (Abu Samaan et al., 2019). Among breast cancer subtypes, paclitaxel is more effective on basal and ERBB2+ subtype compared to luminal and normal like cancers (Rouzier et al., 2005). Apart from the use of paclitaxel for the treatment of different intrinsic subtypes of breast cancer, paclitaxel is being used to treat metastatic breast cancer (Gradishar, 2012). Several studies have been performed to identify the probable biomarkers but none of the identified biomarkers have been applied to clinics yet. Parardiso et al. showed  $\beta$ -tubulin III immunohistochemical expression as a predictive biomarker for paclitaxel in a cohort of 72 patients of advanced breast cancer (Paradiso et al., 2005). Chang et al. proposed 92 gene list capable of classifying breast cancer patients into docetaxel (a derivative of paclitaxel) sensitive and resistant groups in neoadjuvant setting (Chang et al., 2003). Sun et al. showed Phosphoglycerate kinase-1 (PGK1) as a prognostic biomarker for breast cancer treatment with paclitaxel (Sun et al., 2015). Murray et al. summarized different predictive biomarkers for taxane resistance which include Multidrug Resistance Protein 1 (MDR1), P-glycoprotein (Pgp) and  $\beta$ -Tubulin (Murray et al., 2012). Xie et al. showed microtubule binding proteins such as MAP2, Tau, parkin and survivin as biomarkers for paclitaxel response (Xie et al., 2016). Pavlíková et al. compared MCF7 sensitive and resistant cells

to paclitaxel and found that HSP27 and cathepsin were upregulated in sensitive cells and TRIP6 was upregulated in resistant cells (Pavlíková et al., 2015). Xu et al. showed PDCD4 as a biomarker for both paclitaxel and cancer progression (Xu et al., 2015). Hodgkinson et al. summarized different paclitaxel resistant biomarkers in breast cancer cell line among which downregulated markers are cytokeratin 19, HSP27 and phosphoglycerate kinase-1, and upregulated genes included 14-3-3 epsilon, proliferating cell nuclear antigen, sorcin and stathmin (Hodgkinson et al., 2010).

### **1.3.2. Doxorubicin**

Doxorubicin is an important anticancer drug that belongs to anthracycline class of compounds and is used against both solid and liquid tumors. Different models have been presented regarding cell death in response to treatment of doxorubicin which include formation of DNA adducts leading to increased torsional stress at active promoters, trapping topoisomerase II at breakage site, oxidative stress and ceramide overproduction (Yang et al., 2014). Several studies have been proposed to identify doxorubicin sensitivity and resistance biomarkers. AbuHammad et al. concluded that doxorubicin resistant MCF7 breast cancer cell line showed metabolizing genes such as CYP1A1 and CYP1A1 up-regulation along with drug efflux pump genes and other genes such as TOP2A and MET were downregulated (AbuHammad and Zihlif, 2013). Liang et al. showed that a circular RNA KDM4C when upregulated, decreased doxorubicin resistance and tumor progression (Liang et al., 2019). Long et al. showed that microRNA miR-193b reverses doxorubicin in breast cancer by decreasing MCL-1 expression (Long et al., 2015). Piasecka et al. reported a miRNA cluster involvement in determining TNBC cells resistance/sensitivity to doxorubicin (Piasecka et al., 2018). Sun et al. showed that miRNA-574 is responsible for doxorubicin resistance as it downregulates SMAD4 (Sun et al., 2018). Zang et al showed that Notch signaling plays an important part in cancer cells' response to drugs and downregulation of Notch-1 increased breast cancer cells sensitivity to doxorubicin (Zang et al., 2010). Similar results were shown by Kim et al. (Kim et al., 2015). Moreover, TNBC cells' sensitivity to doxorubicin is increased if BCL-2 expression is decreased (Inao et al., 2018). Moriera et al. showed STAT3 as a potential biomarker for doxorubicin resistance in breast CSCs (Moreira et al., 2018). STAT3 importance in CSC drug resistance was also

reported by Cheng et al. which suggested Stat3/Oct-4/c-Myc signal circuit as liable for doxorubicin resistance (Cheng et al., 2018). Smith et al. performed antibody microarrays to determine doxorubicin resistance biomarkers and reported association of decreased expression of tyrosine phosphorylated form of ERK with doxorubicin resistance along with cyclin B1 and cytokeratin 18 (Smith et al., 2006). Munkácsy et al. reported PSMB7 as bad prognostic biomarker of breast cancer patients and its higher level were associated with doxorubicin resistance (Munkácsy et al., 2010). Loi et al. reported role of CD73 as poor prognosis biomarker of breast cancer and resistance to anthracyclines (Loi et al., 2013). Hodgkinson et al summarized several doxorubicin resistant biomarkers in breast cancer cell line among which downregulated markers alpha tropomyosin, cytokeratin 8, cytokeratin 19, peroxiredoxin 2 and septin 2, and upregulated markers included 14-3-3 sigma (stratifin), ATP synthase  $\beta$ , cathepsin D chain B, cyclophilin B and mitotic checkpoint protein BUB 3 (Hodgkinson et al., 2010).

Several compounds have also been reported which are synergistic to doxorubicin and enhance its activity when breast cancer cells are treated with a combination of these drugs. Li et al. showed a gamma secretase inhibitor to increase MDA-MB-231 sensitivity to doxorubicin (Li et al., 2015). Marinello et al. showed that metformin is a potential drug to prevent doxorubicin resistance in breast cancer cell lines (Marinello et al., 2019).

### **1.3.3. Cisplatin**

Cisplatin is a chemotherapeutic drug which belongs with platinum-based compounds used in treatment of cancer. It is used in treatment of several cancers including breast cancer, bladder cancer, head and neck carcinoma, lung cancer, testicular cancer and ovarian cancer. Cisplatin induces apoptosis in cancer cells through crosslinking with purines bases on DNA, thus it causes DNA damage and hampers DNA repair. Its common side effects include kidney issues, allergic reactions, decreased immunity, gastrointestinal problems and hemorrhages (Dasari and Tchounwou, 2014). Silver et al. proposed cisplatin use in neoadjuvant setting for TNBC patients which show decreased BRCA expression (Silver et al., 2010). Byrski et al. reported high pathological response (pCR) in patients harboring BRCA1 mutations and treated with cisplatin (Byrski et al., 2010). Several studies have

previously reported cisplatin resistance/sensitivity biomarkers. Hodgkinson et al. summarized several cisplatin resistant biomarkers in breast cancer cell line among which downregulated markers are beta tubulin type 3, cytokeratin 17, glutathione S transferase mu 3, HSP27, isocitrate dehydrogenase, peroxiredoxin 4, ribosomal protein P0 and tropomyosin alpha, and upregulated markers included electron transfer flavoprotein beta, heterogeneous nuclear ribonucleoprotein A3, hydroxyprostaglandin dehydrogenase-15 (NAD), matrix metalloproteinase 9, peptidyl-prolyl isomerase B and proteasome beta 1 subunit (Hodgkinson et al., 2010). Xu et al. reported ER- $\alpha$  promoter as resistance biomarker of cisplatin in TNBC patients (Xu et al., 2018a). Pyndiah et al. reported c-MYC as promoter of cisplatin resistance in cancer cells (Pyndiah et al., 2011). Chen et al. summarized several miRNAs which play important part in imparting breast cancer cells with cisplatin resistance/sensitivity (Chen et al., 2016). Xie et al. reported mir-519d as cisplatin sensitivity regulator by down regulating MCL1 expression in breast cancer stem cells (Xie et al., 2017). Leong et al reported p63 as good prognosis biomarker in a subset of sporadic TNBC breast cancer and is responsible for cisplatin sensitivity (Leong et al., 2007). Liu et al reported collagen 1A1 as poor prognostic factor for breast cancer metastasis and its upregulation was associated with cisplatin sensitivity, emphasizing its role as a potential drug target for the patients with higher expression of this gene (Liu et al., 2018a). Zhu et al. proposed ER-  $\alpha$ 36 as cisplatin resistance biomarker and this process is mediated through nongenomic estrogen signaling in breast cancer cells (Zhu et al., 2018).

#### **1.4. Breast cancer plasticity**

Apart from EMT and stemness plasticity in breast cancer, hormone receptor biomarker based and PAM50 based subtype switching is observed.

ER, PR and Her2 IHC biomarkers play an important role in decision making for breast cancer patients' treatment plans. Based upon the expression of these hormone receptors either endocrine or chemotherapeutic treatment strategies, are selected in clinics. ER<sup>+</sup> and PR<sup>+</sup> breast cancer subtypes are treated with hormonal therapy for tumors with low risk of recurrence but if the risk of recurrence is high then chemotherapy treatment is also selected. For HER2<sup>+</sup> subtype, trastuzumab or pertuzumab along with chemotherapeutics is used to

treat early breast cancer patients. And TNBC patients are treated with chemotherapeutics (Harbeck et al., 2019). So, it is very important to identify breast cancer subtype specifically and treat patients with an appropriate plan. Unfortunately, the expression of these IHC is not stable in breast cancer. Almost 33% of breast cancer patients express a change in the status of IHC markers and almost one out of seven patients show a change in HER2 status (Lindstrom et al., 2012). Hirata et al. showed that breast cancer patients when treated with neoadjuvant therapy showed a significant change in IHC status. The overall change in status was determined at 16%. And patients who switched their status and no hormonal therapy was used for treatment, showed worse overall and disease free survival (Hirata et al., 2009). Kasami et al. showed a significant difference in PR status before and after neoadjuvant chemotherapy (Kasami et al., 2008). Neubauer et al. showed significant change for all 3 IHC markers and the overall change for this study was 23% (Neubauer et al., 2008). Adams et al. showed a significant increase in HER2 IHC positivity in breast cancer patients before and after neoadjuvant chemotherapy (Adams et al., 2008). Tacca et al. showed 23% switch in hormonal receptors status in either direction before and after neoadjuvant chemotherapy (Tacca et al., 2007). Thompson et al. showed that out of 208 breast cancer patients, ER status switch was observed among 10.2% patients. And this percentage for PR was 24.8% and HER2 was 2.9% (Thompson et al., 2010). Switch in IHC status is also reported in between primary and metastatic breast cancer tissue as well. Sari et al. reported significant difference for hormone receptor between primary and metastatic tumors where the change in ER status was 36%, change in PR status was 54% and for HER2, the switch in status was 14% (Sari et al., 2011). Aurilio et al. performed a meta-analysis in which they observed 24% switch for ER status, 46% switch in PR status and for HER2 13% from negative to positive status between primary and recurrent tumor. Additionally 14% tumors showed gain of ER status, 15% showed gain of PR status and 5% showed gain of HER2 status on recurrence (Aurilio et al., 2014). A number of other studies have validated or showed the same results (Zidan et al., 2005; Broom et al., 2009; Ataseven et al., 2012; Niikura et al., 2012; Dieci et al., 2013; Kinoe et al., 2018; Yang et al., 2020). In literature previously, this switch of hormonal receptor status has been associated with administration of chemotherapy (Anand and Velayudhan, 2016; Erdem et al., 2017; Vogel et al., 2019).

As breast cancer is divided into 5 intrinsic molecular subtypes which are LumA, LumB, Her2, Basal and Normal. In clinics, breast cancer samples are subtyped using PAM50 test which uses the expression of 50 genes to assign each sample to its molecular class (Parker et al., 2009). As opposed to IHC subtyping plasticity, not many studies have been performed to determine if PAM50 subtyping is switched upon chemotherapy treatment or metastasis. Cejalvo et al. analyzed primary and metastatic samples to study PAM50 subtype switching in 123 patients. For these patients, metastasis was recorded for several organs including lymph node, lung, bone, skin and peritoneum among others. Basal showed the highest concordance for switching between primary and metastatic tissue as no switching was observed. HER2 and LumB showed similar switching as the rate of the switch was 23.08% and 30%. The highest switching percentage was observed for LumA at 55.32% and most of these samples switched to LumB and HER2 subtype (Cejalvo et al., 2017).

### **1.5. Metastatic biomarkers for BC**

Breast cancer mortality rate has significantly decreased during the last 20 years. But metastatic breast cancer still holds the major responsibility for breast cancer lethality as only 5% of breast metastatic patients survive long term as this disease have a median overall survival of 20-28 months (Gobbini et al., 2018). Breast cancer metastasis takes place through several steps sequentially which include migration of tumor cells after achieving migration properties towards blood vessels, transfer of these cells to blood supply known as intravasation, survival of tumor cells in blood circulation, extravasation to a different host organ, colonization and metastatic tissue growth with the manifestation of clinical symptoms (Lambert et al., 2017). Breast cancer metastasis is detected through clinical symptoms, biopsies of metastasis, radiological examinations, imaging techniques and metastasis biomarkers (Scully et al., 2012). Breast cancer metastasis is mostly observed for lymph node, brain, lungs and bone (Lorusso and Rüegg, 2012). EMT is reported to play an important role in metastasis as mesenchymal cells show increased migration and resistance to therapy along with increased tumorigenicity (Dongre and Weinberg, 2019). EMT transcription factors (Snail1 and Twist1) overexpression not only give metastatic

properties to expressing cells but also help neighboring cells in achieving metastasis as well (Neelakantan et al., 2017).

In breast cancer, several metastasis biomarkers have been reported previously. Cabioglu et al. showed that CCR7 and CXCR4 are axillary lymph node metastatic biomarkers in early stage breast cancer (Cabioglu et al., 2005). Cai et al. reported TPX2, KIF2C, CDCA8, BUB1B, and CCNA2 as prognostic markers for distant metastasis in breast cancer (Cai et al., 2019b). Merdad et al. showed MMP9 can be used as a biomarker for cancer invasion and metastasis (Merdad et al., 2014). Zurita et al. proposed hypermethylated promoters of genes 14-3-3- $\sigma$  and ESR1 as breast cancer metastasis biomarkers in serum (Zurita et al., 2010). Tseng et al. proposed a biomarker model to diagnose breast cancer metastasis and this model included clinical features such as TNM staging, Age, ER, PR and HER2 along with serum biomarkers (CA15-3, CEA, and sHER2). This proposed model can predict metastasis 90 days ahead of the event (Tseng et al., 2019). Fackler et al. proposed a methylation specific PCR assay for 10 genes which can be used for therapeutic and diagnostic purposes in breast cancer metastasis. Ten genes included in the assay were AKR1B1, COL6A2, GPX7, HIST1H3C, HOXB4, RASGRF2, TM6SF1, ARHGEF7, TMEFF2 and RASSF1 (Fackler et al., 2014). Chen et al. proposed 2 micro-RNAs (miRNA-10b and miRNA-373) as biomarkers for lymph node breast cancer metastasis (Chen et al., 2013b). Eichelser et al. showed increased expression of miR-17 and miR-155 in non-metastatic samples when compared with metastatic samples in the serum of breast cancer patients (Eichelser et al., 2013). Roth et al. showed that the expression of miR-10b, miR-34a, and miR-155 in the blood significantly correlated with overt metastasis (Roth et al., 2010). Thomassen et al. performed a meta-analysis for 1200 patients in total to identify potential metastatic candidate genes. PSD3, FOS, LPL, EPHX2 and DIRAS3 were proposed as metastasis suppressor genes while AURKA, RECQL4, PRMT7 and GINS2 were reported as metastasis promoter genes (Thomassen et al., 2009). Györfy et al. generated a dataset of 1079 patients by combining previously published datasets and identified a gene signature of 376 genes which can predict breast cancer recurrence (Györfy and Schäfer, 2009). Wang et al. identified 76 gene signature which could predict distant metastasis of breast cancer patients with lymph node negative status (Wang et al., 2005). Minn et al. proposed a gene signature related with lung metastasis in breast cancer

patients and they named this “lung metastasis gene expression signature”. This signature included genes such as SPARC, TNC, PTGS2, ANGPTL4, MMP2, FSCN1, MMP1, KRT1B1, CXCL1, EREG, ID1, LY6E and RARRES3 (Minn et al., 2007). Barraclough et al. performed a meta-analysis for breast cancer cohorts and propose dCOX2 and RPM2 as metastasis biomarkers (Bell et al., 2017). GATA3 expression was decreased in metastatic tissue in a subset of ER<sup>+</sup> irrespective of hormonal treatment (Mehra et al., 2005). Using matched primary and metastatic tissues of 16 breast cancer patients, Iwamoto et al. showed that VEGFA and DNMT3A are overexpressed in brain metastasis when compared with primary tissues (Iwamoto et al., 2019). Vargas et al. showed in their meta-analysis that ER $\alpha$  is responsible for increased bone metastasis and this expression is unique to bone metastasis and is not observed for lung metastasis (Vargas et al., 2019). Matrone et al. showed that metastatic breast cancer overexpresses tau which leads to microtubule stabilization providing further support in metastasis (Matrone et al., 2010). Han et al. proposed that NM23 is responsible for decreased lymph node metastasis (Han et al., 2017). Kim et al. correlated increased PD-L1 expression with increased lymph node metastasis (Kim et al., 2017). Li et al. integrated three datasets for meta-analysis and showed that CD80 and ISG15 leads to increased breast cancer progression and metastasis in breast cancer (Li et al., 2020). Pan et al. through a meta-analysis, showed that higher expression of STAB1 is related with increased metastasis leading to worse prognosis in breast cancer (Pan et al., 2016). CXCR4 was shown to be highly associated with breast cancer metastasis and is also expressed on breast cancer stem cells (Zhang et al., 2014).

Additionally, Shriver et al. showed that loss of OBSCN gene is associated with EMT leading to increased tumorigenicity and metastasis (Shriver et al., 2015). In metastatic breast cancer, Imani et al. showed through meta-analysis that TWIST1, ZEB1, SLUG and SNAIL1 IHC expression was associated with worse prognosis (Imani et al., 2016). Chakrabarti et al. showed that ELF5 is responsible for EMT inhibition resulting in metastasis repression through decreased SNAIL2 (Chakrabarti et al., 2012). Yan et al. showed that GATA3 expression results in MET leading to decreased breast cancer metastasis (Yan et al., 2010). Gjerdrum et al. showed that Axl expression is related to cancer survival and, cells who undergo EMT express increased Axl in metastasized tissue (Gjerdrum et al., 2010). DiMeo et al. showed that Wnt signaling in breast cancer, plays a

major role in EMT resulting in metastasis and Wnt signaling inhibition resulted in the loss of mesenchymal properties and metastasis as well (DiMeo et al., 2009). Cai et al. showed that ectopic overexpression of miR-374a promoted both EMT and metastasis *in vitro* and *in vivo* (Cai et al., 2013). Additionally, Yan et al. showed that a decrease in the number of circulating tumor cells is correlated with treatment response (Yan et al., 2017).

Although these studies show different biomarkers for breast cancer metastasis but the limitation of these studies, is the number of patients and metastatic cohorts used. Metastatic biomarkers identified through analyzing multiple cohorts and with an increased number of patients will hold higher importance and will be the most suitable candidates for further validation.

### **1.6. Role of MLANA and INHBA in Melanoma**

Through epidemiological studies, researchers have shown a link between cutaneous melanoma and breast cancer. Carriers of BRCA gene mutations have also a high risk of melanoma and similarly, carriers of CDKNA2 gene have a higher risk of developing breast cancer (Goggins et al., 2004). Similarly, Ho et al. analyzed the Irish Cancer Registry for identifying any correlations between malignant melanoma and breast cancer. And authors found a link between both conditions suggesting that survivors of both cancers should undergo routine check-ups for the second cancer (Ho et al., 2011).

Melanoma shows various phenotypes based upon stemness, EMT and invasion/proliferation. In our previous publication, we showed that these phenotypes are relatively related to each other in melanoma cells. Additionally, using only two genes Inhibin subunit A (INHBA) and Melanin A (MLANA), melanoma cells can be classified in their corresponding classes based upon stemness, EMT and invasion/proliferation. In this study, only two genes were used as a classifier as we wanted to find minimal number of genes for classification (Senses et al., 2017). MLANA is a critical regulator of pigmentation and is previously associated with the proliferative phenotype of melanoma (Arozarena and Wellbrock, 2019). In melanoma, a large proportion of cells express MLANA which also acts as melanoma antigen for T cells. And MLANA is downregulated in hypoxic conditions (EMT rich regions) which suggests that proliferative/epithelial cells

express MLANA and MLANA can be used as epithelial and melanocytic marker (Widmer et al., 2013).

INHBA is overexpressed in the invasive phenotype of melanoma and is related to extra cellular microenvironment modifications, decreased proliferation and high motility(Hoek et al., 2008). INHBA knockdown *in vitro* was associated with decreased EMT (Sinnberg et al., 2018). So, this data shows that MLANA biomarker represents non-stem/epithelial/proliferative phenotype of melanoma and INHBA represents stem/mesenchymal/invasive phenotype. Finally, these conclusions along with similarities of breast cancer and melanoma, warrants the question if stem/non-stem and mesenchymal/epithelial markers in melanoma can be used as melanoma prognostic biomarkers.

## 2. Aim of the study

Breast cancer shows plasticity and switches among its various subtypes depending upon various factors such as treatment, metastasis and microenvironment. This plasticity has been reported to occur through EMT which in turn, has been associated with stemness. It is likely that when breast tumors are treated with traditional therapy, breast cancer stem cells are not killed; or due to EMT, mesenchymal cells are generated which are resistant to therapy. In this study, we aimed to identify a robust gene list that could classify BC samples regarding their stemness and EMT characteristics. Using this gene list, we aimed to determine whether a thus generated categorization can be applied to all BC cohorts, and whether this could be used to determine drug sensitivity and prognosis. We also aimed to generate a classification of BC based on the same gene list, but via continuous variables. We used this second approach to test if we could trace phenotype plasticity in BC, and how this related to chemosensitivity. We asked if our gene list had any prognostic importance, which could lead it to be used in the clinics.

Secondly, we aim to identify specific biomarkers for paclitaxel, cisplatin and doxorubicin sensitivity, as these drugs are routinely used to treat breast cancer patients. By identifying such biomarkers, we hypothesized that these treatments can be personalized to patients which would show a higher response to these drugs.

Thirdly, we aimed to highlight breast cancer plasticity for both IHC based (ER, PR and HER2) and PAM50 subtypes. In breast cancer, patients receive treatment based on their IHC marker expression and subtype classifications. Therefore, if subtype changes occur upon treatment or metastasis, the treatment plan should be revised according to the new subtype to better achieve the prognostic outcome. We also aimed to evaluate metastatic biomarkers in breast cancer using multiple cohorts *in silico*.

Lastly, as we had previously shown two markers, Inhibin (INHBA) and Melanin A (MLANA), to subtype melanoma tumors into invasive/proliferative, epithelial/mesenchymal and stem/non-stem phenotypes similar to our stem/non-stem and

epithelial/mesenchymal classification in breast cancer, we aimed to study the prognostic significance of INHBA and MLANA in comparison to our breast cancer gene list.



### 3. Materials

#### 3.1. General laboratory reagents

Reagents and chemicals for general use in the laboratory, were bought from several sources as shown in Table 3.1.

**Table 3.1: General laboratory reagents and vendors**

| Reagent / Chemical                        | Company                                         |
|-------------------------------------------|-------------------------------------------------|
| Ethanol                                   | SIGMA-ALDRICH (St. Louis, MO, USA)              |
| Isopropanol                               | SIGMA-ALDRICH (St. Louis, MO, USA)              |
| Chloroform                                | SIGMA-ALDRICH (St. Louis, MO, USA)              |
| DMSO (Dimethyl sulfoxide)                 | AppliChem (Darmstadt, Germany)                  |
| RNase Zap                                 | Ambion (Carlsbad, CA, U.S.)                     |
| Nuclease Free Water                       | Ambion (Carlsbad, CA, U.S.)                     |
| Qubit™ RNA BR Assay Kit                   | Invitrogen (Carlsbad, CA, USA)                  |
| DNA-free™ Kit                             | Invitrogen (Carlsbad, CA, USA)                  |
| RevertAid First Strand cDNA Synthesis Kit | Thermo Scientific (Waltham, MA, USA)            |
| OneTaq Quick-Load 2X Master Mix           | New England Biolabs (Ipswich, MA, USA)          |
| iTaq Universal SYBR Green Supermix        | Bio-Rad Laboratories, Inc., (Berkeley, CA, USA) |

50X TRIS acetate EDTA was prepared by dissolving 242 g of Tris Base, 57.1 ml of glacial acetic acid and 37.2 g of Tritiplex III (EDTA) in 1 liter of dd H<sub>2</sub>O.

#### 3.2. General laboratory equipment

Thermo Scientific™ NanoDrop™ One Microvolume UV-Vis Spectrophotometer (Thermo Scientific (Waltham, MA, USA) and Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) were used to check the quality of RNA samples before and after DNase treatment respectively. Gel electrophoresis was performed in Gel electrophoresis system (Hoefer Inc., Holliston, MA). For luminescence-based assays, Synergy HTX multi-mode reader (Biotek, VT, USA) was used.

### 3.3. Cell culture reagents

Cell culture media and reagents were procured from different sources as shown in Table 3.2.

**Table 3.2: Cell culture media and reagents**

| Reagents                                        | Company                              |
|-------------------------------------------------|--------------------------------------|
| Dulbecco's Modified Eagle Medium (DMEM)         | Lonza (Basel, Switzerland)           |
| Rosswell Park Memorial Institute Medium (RPMI)  | GIBCO (Carlsbad, CA, USA)            |
| B27 spheroid media supplement                   | Invitrogen, Carlsbad, CA, USA        |
| Epidermal Growth Factor human (EGF)             | SIGMA-ALDRICH (St. Louis, MO, USA)   |
| Fibroblast Growth Factor-Basic human (FGF)      | SIGMA-ALDRICH (St. Louis, MO, USA)   |
| Heparin sodium salt                             | SIGMA-ALDRICH (St. Louis, MO, USA)   |
| Trypsin-EDTA                                    | SIGMA-ALDRICH (St. Louis, MO, USA)   |
| Fetal Bovine Serum (FBS)                        | GIBCO (Carlsbad, CA, USA)            |
| L-Glutamine                                     | Lonza Group (Switzerland)            |
| Fetal Bovine Serum (FBS)                        | Lonza Group (Switzerland)            |
| Penicillin –Streptomycin                        | HyClone (Rockford, USA)              |
| CellTiter-Glo® Luminescent Cell Viability Assay | Promega (Fitchburg, Wisconsin, USA)  |
| TrypLE™ Select Enzyme (1X), no phenol red       | Life Technologies (Waltham, MA, USA) |
| TRIzol® Reagent                                 | Life Technologies (Waltham, MA, USA) |

### 3.4. Cell culture media and solutions

Breast cancer cell lines BT474, BT20, CAL51, JIMT1, MCF7, MDA-MB-157, MDA-MB-231, MDA-MB-453, MDA-MB-468, MDA-MB-436, ZR751, were cultured in DMEM media whereas HCC38, HCC70, HCC202, HCC1143, HCC1937, HCC1954 and T47D were cultured in RPMI media.

To prepare cell culture media for 2D cell culture, 10% FBS, 1% L-Glutamine and 1% Penicillin –Streptomycin of total media, were filtered using 0.2 µm syringe filters and added to respective medium to make complete medium.

To prepare cell culture media for 3D cell culture (mammosphere culture), 1% L-Glutamine and 1% Penicillin –Streptomycin of total media, were filtered using 0.2 µm syringe filters and added to respective medium in addition to 1x B27, EGF 10ng/ml, FGF 20ng/ml and heparin 2 µg/ml to make complete medium. Once prepared medium was used for 7 days after which remaining medium was discarded and new medium was prepared.

10X Phosphate buffer saline (PBS) was prepared by dissolving 80g of NaCl, 2g of KCl, 14.4g of Na<sub>2</sub>HPO<sub>4</sub> and 2.4g of KH<sub>2</sub>PO<sub>4</sub> in 800 ml of ddH<sub>2</sub>O. pH of the solution was adjusted to 7.4 and volume was increased to 1 liter by adding more ddH<sub>2</sub>O. 100ml of prepared 10X PBS was added to 900ml of ddH<sub>2</sub>O generating 1X PBS. 1X PBS was then autoclaved and filtered before use in cell culture.

To freeze cells in liquid nitrogen, 10ml DMSO was added to 90ml FBS to generate freezing medium.

For cell cytotoxicity analysis Lapatinib was bought from CAYMAN and Midostaurin from Sigma Aldrich. Both drugs were dissolved in DMSO generating 50mM and 5mM stocks respectively.

## **4. Methods**

### **4.1. Cell culture conditions**

All cell lines were cultured in their respective medium (as explained in Materials section 3.4) at 5% CO<sub>2</sub> incubator at 37 °C.

### **4.2. Culturing cryopreserved cells**

Cryovials containing cell lines were stored in liquid nitrogen. To start a culture, cryovials were removed and thawed immediately in a water bath at 37 °C. Cells from these vials were transferred then to 15ml falcon tubes. 5 ml medium was added to cells drop by drop. Then falcon tubes were centrifuged at 1500 rpm for 3 minutes. The supernatant from each falcon was removed and the cell pellet was dissolved in 6 ml of complete medium. These cells were then transferred to 25 cm<sup>2</sup> flask and kept at 5% CO<sub>2</sub> incubator at 37 °C.

### **4.3. Cell culture**

Cell lines after achieving more than 70% confluency were passaged. For passaging, the medium was aspirated and cells were washed with 5ml of 1X PBS. Then trypsinization was performed. 1-2 ml Trypsin-EDTA was added to each flask and then flasks were placed in an incubator at 37 °C for 2-3 minutes. Upon detachment of cells, these were collected in 5ml of the medium. This cell suspension was centrifuged in a falcon tube at 1500 rpm for 3 minutes. After centrifugation, the supernatant was removed and the cell pellet was resuspended in 3 ml medium. 1 ml of this suspension was transferred to 75 cm<sup>2</sup> flask containing 8 ml complete medium.

### **4.4. Cell cryopreservation**

To cryopreserve cells in liquid nitrogen, cells grown in a flask were washed with 1X PBS and then trypsinized. 5 ml medium was added to those trypsinized cells and this mixture was centrifuged at 1500 rpm for 3 minutes. The supernatant was removed and cell pellet was resuspended in freezing media (explained in Materials section 3.4). This cell suspension was then transferred to cryovials. These vials were then labeled and stored at -20 °C for 2 hours and then moved to liquid nitrogen.

#### **4.5. Mammosphere culture (3D cell culture)**

Cells were cultured initially in 2D conditions and upon reaching 80-90% confluency, these were trypsinized and centrifuged at 800rpm for 4 minutes. The supernatant was removed and cell pellet was resuspended in 8 ml of sphere media. This cell suspension was then transferred to 75 cm<sup>2</sup> ultra-low attachment flasks. These flasks were placed in an incubator for 3 days. Cells formed spheres after 3 days. Spheres were separated from single cells using 40 µm strainer. These spheres were then washed and resuspended in PBS. Spheres were then centrifuged at 800 rpm for 4 minutes. The supernatant was removed and 3 ml TrypLE was added to the cell pellet. The mixture was pipetted rigorously to dissociate mammospheres both physically and then was placed into an incubator. Then 6 ml of 1X PBS was added and the whole mixture was centrifuged again at 800 rpm for 4 minutes. The supernatant was removed and cells were suspended in 8 ml of media and transferred to 75 cm<sup>2</sup> ultra-low attachment flask. This procedure was repeated for 6 passages of mammospheres.

#### **4.6. RNA extraction**

To extract RNA from cells grown in adherent (2D cell culture condition), 80% confluent 75cm<sup>2</sup> flask was washed with 1x PBS. Then 6 ml ice cold PBS was added and using cell scraper, cells were scrapped. These cells were then collected and transferred to falcon tubes which was followed with centrifugation at 1500 rpm for 4 minutes at 4 °C. The supernatant was removed and the cell pellet was stored at -80 °C.

To extract RNA from mammospheres, spheroids were centrifuged at 1500 rpm for 4 minutes. The supernatant was removed and the cell pellet was resuspended in 6 ml ice cold 1X PBS followed by centrifugation. At 1500 rpm for 4 minutes. Then the supernatant was removed and the cell pellet was stored at -80 °C.

To extract RNA from cell pellet, 1 ml TRIzol reagent was added and samples were transferred to separate Eppendorf followed by 5 minutes incubation at room temperature. 200 µl chloroform was added to each Eppendorf and pipetted vigorously followed by 10 minutes incubation at room temperature. Samples were then centrifuged at 13000 rpm for

15 minutes at 4 °C. Upper phases from each tube which contained RNA, were then transferred to a separate tube. 500 µl isopropanol was added to each sample followed by 10 minutes incubation at room temperature and 13000 rpm centrifugation for 15 minutes at 4 °C. Then the supernatant was removed and the pellet was washed with 1 ml of ethyl alcohol. Pellets were then incubated for 10 minutes incubation at room temperature followed by 13000 rpm centrifugation for 15 minutes at 4 °C. Then ethyl alcohol was removed and the pellets were dried. These pellets were then dissolved in Nuclease free water. Each sample was incubated for 15 minutes in the Thermolyne heater. RNA quantity and quality in each sample was measured by Nanodrop spectrophotometer ND-1000.

#### **4.7. DNase treatment**

To remove DNA contamination in the samples, these were treated DNA Free kit. Each sample was diluted to 200 ng/µl. 44 µl of the sample was mixed with 1 µl DNase I and 5 µl DNase buffer. These samples were incubated for 30 minutes at 37 °C. The reaction was terminated with 5 µl of DNase inactivating reagent with vigorous pipetting. Then samples were centrifuged at 1000 rpm for 2 minutes. The supernatant contained DNA free RNA and was transferred to a newly labeled eppendorf.

#### **4.8. RNA quantification**

Qubit fluorometer was used to measure the quantity RNA using Qubit RNA BR assay. 1 µl Qubit BR reagent was added to 199 µl buffer to prepare a working solution for each sample. Calibration of the instrument was performed using standards that were prepared by adding 10 µl of the standard to 190 µl of working solution. After calibration, 2 µl of each sample was added to 198 µl of working solution. Measurements were taken after 2 minute incubation at room temperature.

#### **4.9. cDNA synthesis**

For each sample, 500 ng of total RNA was used to reverse transcribe into cDNAs using RevertAid First Strand cDNA synthesis Kit. Random hexamer primers were used for this reaction. 1 µl hexamer primer were added to 11 µl of RNA and water mixture containing 500ng RNA. This mixture was incubated for 5 minutes at 65 °C. Samples were spun and

placed on ice. Reaction master mix was prepared with 1 µl of 5x reaction buffer, 1 µl of Riboblock RNase inhibitor, 2 µl of 10mM dNTP master mix and 1 µl of RevertAid RT for each sample. To each sample, 8 µl of master mix was added to complete 20 µl reaction. - RT controls were prepared using same mixture except reverse transcriptase was swapped with nuclease free water to check genomic DNA contamination. A no template control was also prepared where nuclease free water was used instead of sample RNA to check reagent contamination. All these samples were then incubated at 25 °C for 10 minutes, 42°C for 60 minutes followed by reaction termination at 70°C for 5 minutes. All these samples were stored at -80°C.

#### **4.10. End point polymerase chain reaction**

End point PCR reaction was performed to check if produced cDNAs were free of genomic contamination. All samples including control were amplified with 18S RNA primers (forward primer: 5'-CGTGCATTTATCAGATCAAAACCAACC-3' and reverse primer: 5'-ATGGTAGGCACGGCGACTAC-3'). OneTaq™ Quick Load® Master Mix (New England BioLabs) was used. Each reaction included 5 µl of 2X OneTaq master mix, 0.2 µl of forward and reverse primer each, 3.4 µl of nuclease free water and 1.2 µl of cDNA sample. Controls reaction were prepared with the same concentrations except nuclease free water was used instead of sample. Initial denaturation for this PCR was at 94°C for 5 minutes which was followed by 30 cycles (denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 30 seconds). The final extension was at 72°C for 5 minutes.

#### **4.11. Agarose gel electrophoresis**

To detect genomic and reagent contaminations, Agarose gel electrophoresis was performed. 1% agarose gel solution was prepared in 1X TAE buffer via heating in a microwave oven. 5 µl EtBr was added to 100 ml agarose. Agarose solution was then poured in a gel tray and comb was placed. Upon gel solidification, the tray was placed in a tank filled with TAE buffer and samples (8 µl of PCR product mixed with 2 µl of loading dye) along with the DNA ladder (#SM0373 Thermo Scientific Waltham, MA, USA) were

loaded in wells generated from removing comb. Gel was run at 110 volt for 30 minutes and visualized under UV.

#### **4.12. Quantitative Real time Polymerase Chain Reaction (qRT-PCR)**

qRT-PCR was performed using Light Cycler 480 (Roche) with iTaq Universal SYBR Green Supermix (Bio-Rad). Primers were generated as described previously in the Master's thesis (Akbar, 2014). Each reaction was performed in triplicates with no template control for each plate. Reaction conditions were Polymerase activation and DNA denaturation at 95 °C for 30 seconds followed by 40 cycles (denaturation: 15 seconds at 95°C, annealing/extension: 60 seconds at 60°C) followed by melt curve stage 65-95°C (0.5°C increments at 2 seconds/step). GAPDH was used as endogenous reference control and all data was calculated using the  $\Delta\Delta CT$  method (Livak and Schmittgen, 2001).

#### **4.13. Drug cytotoxicity**

Drug cytotoxicity assays were performed both in 2D and 3D cell culture conditions using both CellTiter-Glo® Luminescent Cell Viability Assay Kit (Promega) and CyQUANT Cell Proliferation Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. For 2D cell culture, cells were counted using a hemocytometer and 3000 cells were cultured in each well of 96 well plate. After 24 hours, these cells were treated with drugs at 10 different concentrations (Lapatinib: 50  $\mu$ M-0.001  $\mu$ M, Midostaurin: 5 $\mu$ M-0.0001  $\mu$ M). 72 hours post drug treatment, percent cell viability and cytotoxicity were calculated using both assays and dose inhibition curves were drawn as described previously in the Master's thesis (Akbar, 2014).

Cell viability and cytotoxicity assays were performed for 3D cell culture the same way as 2D assays except 96 well ultra-low attachment plates were used and cells were cultured in mammosphere media.

#### **4.14. siRNA Knockdown and Cell viability**

Using commercially available siRNAs at a concentration of 20nM, ZEB1 (D-006564-04, Dharmacon) and SNAI2 (D-017386-02, Dharmacon) were knocked down along with

control siRNA (SI03650318, Qiagen). Cells were seeded at a density of  $2 \times 10^5$  in 6 well plate in complete medium. When cell confluency reached around 50-60%, complete medium was replaced with medium without antibiotics and Lipofactamine 2000<sup>TM</sup>-siRNA mixture was added. 48 hours after transfection, cells were collected and change in transcript levels was determined using qRT PCR. For cytotoxicity experiments, 3000 cells/well were seeded in 96 well plates. After 24 hours, cells were transfected with siZEB1, siSNAI2 and siControl. Cells were treated with 4 concentrations of 2 drugs (Midostaurin: 0.1, 0.5, 1 and 5  $\mu$ M and Lapatinib: 1, 5, 10 and 50  $\mu$ M) after 24 hours. Cell viability was determined 72 hours after drug treatment, with CellTiter-Glo<sup>®</sup> Luminescent Cell Viability Assay Kit. All treatments were performed in quadruplicates.

#### **4.15. *In silico* analyses**

For *in silico* analyses, all datasets were downloaded from ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) and Gene expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). One selection parameter was that these datasets were generated using Affymetrix and Illumina microarray platform. Using BRB Array Tool, each dataset was RMA normalized. GSE73526 dataset was generated using RNA next generation sequencing platform and for this dataset normalized FPKM values were used. Data was clustered hierarchically using Cluster tree 3.0 program and Java Treeview program was used to draw heatmaps (de Hoon et al., 2004; Saldanha, 2004). GSEA (Gene set enrichment analysis) was performed using according to Broad institute guidelines (Subramanian et al., 2005). Gene ontology database “C5-all” was used in each GSEA analysis. PAM50 signature was used to classify breast cancer samples into 5 molecular classes as described previously: Luminal A (LumA), Luminal B (LumB), HER2 enriched (HER2), Basal and Normal like (Normal) (Parker et al., 2009).

#### **4.16. Statistical analysis**

Different groups were compared with each other using student t.test. GraphPad Prism v 6.01 (GraphPad Software Inc.) was used to generate graphs. Cox regression analysis and Log rank test were performed using “Survival” package in R (Therneau and Grambsch,

2000; Team, 2013). Kaplan Meier analysis was performed using IBM SPSS Statistics v.19 and GraphPad Prism v 6.01.



## 5. Results

### 5.1. Classification of breast cancer based on stemness and EMT

#### 5.1.1. Robustness of the CSC/non-CSC gene list (CNCL) classification

In our previous work, we identified 15 genes (Table 5.1) which could identify cancer stem cells (CSC) from non-CSC cells (Akbar, 2014). Hierarchical clustering of CCLE (GSE36139) (Barretina et al., 2012) and CGP (E-MTAB-783) (Garnett et al., 2012) datasets showed that our genes were differentially expressed in some breast cancer cell lines which could thus be categorized as CSC like and non-CSC like cells.

**Table 5.1: CSC/non-CSC gene list (CNCL)** Table retrieved from (Akbar et al., 2020a).

| Gene Symbol | Gene Name                                       | Behavior in CSC like cells |
|-------------|-------------------------------------------------|----------------------------|
| IRF6        | Interferon Regulatory Factor 6                  | Downregulated              |
| ST14        | Suppression of Tumorigenicity 14                | Downregulated              |
| CDH1        | Cadherin 1                                      | Downregulated              |
| BSPRY       | B-Box and SPRY Domain Containing protein        | Downregulated              |
| CLDN4       | Claudin 4                                       | Downregulated              |
| AP1M2       | Adaptor-related Protein Complex 1, mu 2 Subunit | Downregulated              |
| ZNF165      | Zinc finger protein 165 (CT gene)               | Downregulated              |
| PVRL3       | Poliovirus Receptor-Related 3                   | Upregulated                |
| SLIT2       | Slit Homolog 2                                  | Upregulated                |
| BNC2        | Basonuclin 2                                    | Upregulated                |
| DDR2        | Discoidin Domain Receptor Tyrosine              | Upregulated                |
| VIM         | Vimentin                                        | Upregulated                |
| TMEM158     | Transmembrane Protein 158                       | Upregulated                |
| FN1         | Fibronectin 1                                   | Upregulated                |
| DKK3        | Dickkopf WNT Signaling Pathway Inhibitor 3      | Upregulated                |

To determine if CNCL could cluster the same breast cancer cell lines into same groups across when it was applied to different datasets and platforms, 3 more datasets (GSE24717 (Shats et al., 2011) and GSE50811 (Dezső et al., 2014) (microarray) and GSE73526 (Marcotte et al., 2016) (RNAseq.) were analyzed. Out of all cell lines, only 1 cell line (HCC1419) was misclassified while all other cell lines, present in more than one dataset,

were consistently classified in one of the two groups (CS/M or NS/E) as shown in Table 5.2.

**Table 5.2: CNCL classification robustness across different datasets and platforms.** N/A shows that the cell line was not present in that dataset. Table retrieved from (Akbar et al., 2020a).

| Cell lines | CCL  | CGP  | GSE73526 | GSE50811 | GSE24717 |
|------------|------|------|----------|----------|----------|
| BT549      | CS/M | CS/M | CS/M     | CS/M     | CS/M     |
| Hs578T     | CS/M | CS/M | CS/M     | CS/M     | CS/M     |
| MDAMB231   | CS/M | CS/M | CS/M     | CS/M     | CS/M     |
| CAL51      | CS/M | CS/M | CS/M     | CS/M     | NA       |
| MDAMB157   | CS/M | CS/M | CS/M     | NA       | CS/M     |
| CAL120     | CS/M | CS/M | CS/M     | NA       | NA       |
| HCC1395    | CS/M | CS/M | CS/M     | NA       | NA       |
| MDAMB436   | CS/M | NA   | CS/M     | CS/M     | NA       |
| HBL100     | NA   | NA   | CS/M     | NA       | NA       |
| Hs274T     | CS/M | NA   | NA       | NA       | NA       |
| Hs281T     | CS/M | NA   | NA       | NA       | NA       |
| Hs343T     | CS/M | NA   | NA       | NA       | NA       |
| Hs606T     | CS/M | NA   | NA       | NA       | NA       |
| Hs739T     | CS/M | NA   | NA       | NA       | NA       |
| Hs742T     | CS/M | NA   | NA       | NA       | NA       |
| SKBR7      | NA   | NA   | CS/M     | NA       | NA       |
| SUM1315    | NA   | NA   | CS/M     | NA       | NA       |
| SUM159     | NA   | NA   | CS/M     | NA       | NA       |
| BT20       | NS/E | NS/E | NS/E     | NS/E     | NS/E     |
| HCC1806    | NS/E | NS/E | NS/E     | NS/E     | NS/E     |
| MCF7       | NS/E | NS/E | NS/E     | NS/E     | NS/E     |
| T47D       | NS/E | NS/E | NS/E     | NS/E     | NS/E     |
| BT474      | NS/E | NA   | NS/E     | NS/E     | NS/E     |
| DU4475     | NS/E | NS/E | NS/E     | NS/E     | NA       |
| HCC1143    | NS/E | NS/E | NS/E     | NS/E     | NA       |
| HCC1428    | NS/E | NA   | NS/E     | NS/E     | NS/E     |
| HCC1937    | NS/E | NS/E | NS/E     | NS/E     | NA       |
| HCC1954    | NS/E | NS/E | NS/E     | NS/E     | NA       |
| HCC2218    | NS/E | NS/E | NS/E     | NS/E     | NA       |
| HCC38      | NS/E | NS/E | NS/E     | NS/E     | NA       |
| HCC70      | NS/E | NS/E | NS/E     | NS/E     | NA       |
| MDAMB361   | NS/E | NS/E | NS/E     | NA       | NS/E     |
| MDAMB453   | NS/E | NS/E | NS/E     | NS/E     | NA       |
| MDAMB468   | NS/E | NS/E | NS/E     | NS/E     | NA       |
| SKBR3      | NS/E | NA   | NS/E     | NS/E     | NS/E     |
| UACC812    | NS/E | NS/E | NS/E     | NS/E     | NA       |
| AU565      | NS/E | NA   | NS/E     | NS/E     | NA       |
| BT483      | NS/E | NA   | NS/E     | NA       | NS/E     |

|             |      |      |      |      |    |
|-------------|------|------|------|------|----|
| CAL148      | NS/E | NS/E | NS/E | NA   | NA |
| CAL851      | NS/E | NS/E | NS/E | NA   | NA |
| CAMA1       | NS/E | NS/E | NS/E | NA   | NA |
| EFM19       | NS/E | NS/E | NS/E | NA   | NA |
| HCC1187     | NS/E | NS/E | NS/E | NA   | NA |
| HCC1419     | NS/E | NS/E | NS/E | CS/M | NA |
| HCC1500     | NS/E | NA   | NS/E | NS/E | NA |
| HCC1599     | NS/E | NS/E | NS/E | NA   | NA |
| MDAMB134VI  | NS/E | NS/E | NS/E | NA   | NA |
| MDAMB175VII | NS/E | NS/E | NS/E | NA   | NA |
| MDAMB415    | NS/E | NS/E | NS/E | NA   | NA |
| UACC893     | NS/E | NA   | NS/E | NS/E | NA |
| ZR7530      | NS/E | NS/E | NS/E | NA   | NA |
| EFM192A     | NS/E | NA   | NS/E | NA   | NA |
| EVSAT       | NA   | NS/E | NS/E | NA   | NA |
| HCC1569     | NS/E | NA   | NS/E | NA   | NA |
| HCC202      | NS/E | NA   | NS/E | NA   | NA |
| HCC2157     | NS/E | NS/E | NA   | NA   | NA |
| HDQP1       | NS/E | NA   | NS/E | NA   | NA |
| JIMT1       | NS/E | NA   | NS/E | NA   | NA |
| KPL1        | NS/E | NA   | NS/E | NA   | NA |
| MFM223      | NA   | NS/E | NS/E | NA   | NA |
| OCUBM       | NA   | NS/E | NS/E | NA   | NA |
| ZR751       | NS/E | NA   | NS/E | NA   | NA |
| 184A1       | NA   | NA   | NS/E | NA   | NA |
| 184B5       | NA   | NA   | NS/E | NA   | NA |
| 600MPE      | NA   | NA   | NS/E | NA   | NA |
| COLO824     | NA   | NS/E | NA   | NA   | NA |
| HCC1008     | NA   | NA   | NS/E | NA   | NA |
| HCC2185     | NA   | NA   | NS/E | NA   | NA |
| HCC2688     | NA   | NA   | NS/E | NA   | NA |
| HCC3153     | NA   | NA   | NS/E | NA   | NA |
| HCC712      | NA   | NA   | NS/E | NA   | NA |
| LY2         | NA   | NA   | NS/E | NA   | NA |
| MACLS2      | NA   | NA   | NS/E | NA   | NA |
| MB157       | NA   | NA   | NS/E | NA   | NA |
| MCF10A      | NA   | NA   | NS/E | NA   | NA |
| MCF12A      | NA   | NA   | NS/E | NA   | NA |
| MDAMB330    | NA   | NA   | NS/E | NA   | NA |
| MRKnu1      | NA   | NS/E | NA   | NA   | NA |
| MX1         | NA   | NA   | NS/E | NA   | NA |
| SKBR5       | NA   | NA   | NS/E | NA   | NA |
| SUM102      | NA   | NA   | NS/E | NA   | NA |
| SUM149      | NA   | NA   | NS/E | NA   | NA |
| SUM185      | NA   | NA   | NS/E | NA   | NA |
| SUM190      | NA   | NA   | NS/E | NA   | NA |

|          |      |    |      |      |      |
|----------|------|----|------|------|------|
| SUM225   | NA   | NA | NS/E | NA   | NA   |
| SUM229   | NA   | NA | NS/E | NA   | NA   |
| SUM44    | NA   | NA | NS/E | NA   | NA   |
| SUM52    | NA   | NA | NS/E | NA   | NA   |
| SW527    | NA   | NA | NS/E | NA   | NA   |
| UACC3199 | NA   | NA | NS/E | NA   | NA   |
| YMB1     | NS/E | NA | NA   | NA   | NA   |
| ZR75     | NA   | NA | NA   | NA   | NS/E |
| ZR75B    | NA   | NA | NS/E | NA   | NA   |
| ZRT      | NA   | NA | NA   | NS/E | NA   |

This analysis clearly shows that CNCL is robust in clustering cell lines into CSC like and non-CSC like groups across different platforms and datasets.

#### **5.1.2. Altered Molecular Mechanisms as indicated by differentially expressed genes between CSC and non-CSC like cell lines**

As the breast cancer cell line classification by CNCL was robust, next we aimed to identify altered molecular mechanisms identified by GSEA analysis between CSC and non-CSC like clusters, in CCLE and CGP datasets separately. Genesets enriched in both phenotypes were sorted by significance and were ranked for both CCLE and CGP separately. Geneset ranks from both datasets were added generating a rank sum value to identify commonly enriched genesets. In CSC like cells, genesets related to mesenchymal phenotype such as “Extracellular Matrix”, Basement Membrane” and “Cell Migration” were enriched as shown in Table 5.3.

**Table 5.3: Gene sets upregulated in CSC like (CS/M) cell lines** Table retrieved from (Akbar et al., 2020a).

| Selected Upregulated Genesets                                            | CCL Rank | CGP Rank | Combined Rank |
|--------------------------------------------------------------------------|----------|----------|---------------|
| PROTEINACEOUS_EXTRACELLULAR_MATRIX                                       | 1        | 1        | 2             |
| EXTRACELLULAR_MATRIX                                                     | 2        | 2        | 4             |
| BASEMENT_MEMBRANE                                                        | 5        | 5        | 10            |
| COLLAGEN                                                                 | 7        | 6        | 13            |
| METALLOPEPTIDASE_ACTIVITY                                                | 12       | 7        | 19            |
| BASAL_LAMINA                                                             | 10       | 15       | 25            |
| SKELETAL_DEVELOPMENT                                                     | 11       | 17       | 28            |
| MUSCLE_DEVELOPMENT                                                       | 9        | 24       | 33            |
| CELL_MIGRATION                                                           | 21       | 14       | 35            |
| SULFUR_METABOLIC_PROCESS                                                 | 31       | 8        | 39            |
| TRANSFORMING_GROWTH_FACTOR_BETA_RECEPTOR_SIGNALING_PATHWAY               | 33       | 12       | 45            |
| POSITIVE_REGULATION_OF_RESPONSE_TO_STIMULUS                              | 35       | 11       | 46            |
| TRANSMEMBRANE_RECEPTOR_PROTEIN_SERINE_THREONINE_KINASE_SIGNALING_PATHWAY | 29       | 20       | 49            |
| CELL_CYCLE_ARREST_GO_0007050                                             | 34       | 16       | 50            |
| VASCULATURE_DEVELOPMENT                                                  | 18       | 34       | 52            |
| AXON_GUIDANCE                                                            | 22       | 33       | 55            |
| REGULATION_OF_CELL_MIGRATION                                             | 39       | 28       | 67            |
| REGULATION_OF_CELL_GROWTH                                                | 49       | 23       | 72            |
| METALLOENDOPEPTIDASE_ACTIVITY                                            | 30       | 64       | 94            |
| REGULATION_OF_G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY       | 74       | 35       | 109           |
| PEPTIDYL_TYROSINE_MODIFICATION                                           | 71       | 44       | 115           |

In contrast, in the non-CSC like phenotype, genesets related to epithelial phenotype such as “Tight Junction” were enriched as shown in Table 5.4.

**Table 5.4: Gene sets upregulated in Non-CSC like (NS/E) cell lines.** Table retrieved from (Akbar et al., 2020a).

| Selected Downregulated Gene sets                                                              | CCLE Rank | CGP Rank | Combined Rank |
|-----------------------------------------------------------------------------------------------|-----------|----------|---------------|
| TIGHT_JUNCTION                                                                                | 1         | 1        | 2             |
| APICAL_JUNCTION_COMPLEX                                                                       | 3         | 3        | 6             |
| PROTEIN_BINDING_BRIDGING                                                                      | 10        | 16       | 26            |
| POTASSIUM_CHANNEL_ACTIVITY                                                                    | 15        | 15       | 30            |
| N_ACETYLTRANSFERASE_ACTIVITY                                                                  | 28        | 10       | 38            |
| RESPONSE_TO_HORMONE_STIMULUS                                                                  | 56        | 7        | 63            |
| MICROBODY                                                                                     | 19        | 78       | 97            |
| KINASE_REGULATOR_ACTIVITY                                                                     | 64        | 65       | 129           |
| APICOLATERAL_PLASMA_MEMBRANE                                                                  | 2         | 4        | 6             |
| CALCIUM_INDEPENDENT_CELL_CELL_ADHESION                                                        | 4         | 2        | 6             |
| INTERCELLULAR_JUNCTION                                                                        | 6         | 5        | 11            |
| CELL_JUNCTION                                                                                 | 17        | 13       | 30            |
| ESTABLISHMENT_AND_OR_MAINTENANCE_OF_CELL_POLARITY                                             | 13        | 20       | 33            |
| HYDROLASE_ACTIVITY_ACTING_ON_CARBON_NITROGEN_NOT_PEPTIDE BONDS IN LINEAR AMIDES               | 5         | 37       | 42            |
| REGULATION_OF_MAPKKK_CASCADE                                                                  | 29        | 14       | 43            |
| POTASSIUM_ION_TRANSPORT                                                                       | 24        | 21       | 45            |
| N_ACYLTRANSFERASE_ACTIVITY                                                                    | 33        | 12       | 45            |
| OXIDOREDUCTASE_ACTIVITY_ACTING_ON_THE_ALDEHYDE_OR_OXO_GROUP_OF_DONORS NAD OR NADP AS ACCEPTOR | 11        | 41       | 52            |
| TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_SIGNALING_PATHWAY                              | 25        | 29       | 54            |
| RESPONSE_TO_BACTERIUM                                                                         | 50        | 6        | 56            |
| OXIDOREDUCTASE_ACTIVITY_ACTING_ON_THE_ALDEHYDE_OR_OXO_GROUP_OF_DONORS                         | 20        | 51       | 71            |

This analysis showed that our CSC phenotype has mesenchymal features (therefore, from here on named as CS/M) and the non-CSC phenotype features epithelial characteristics (from here on named as NS/E). We validated this observation by analyzing E-cadherin (E-cad / CDH1, an epithelial marker) protein expression *in silico* in 2 different datasets (TCPA

and Marcotte et al. GSE73526). As shown in Figure 5.1, E-cad expression was consistently and significantly higher in the non-CSC (CS/M) cell lines.

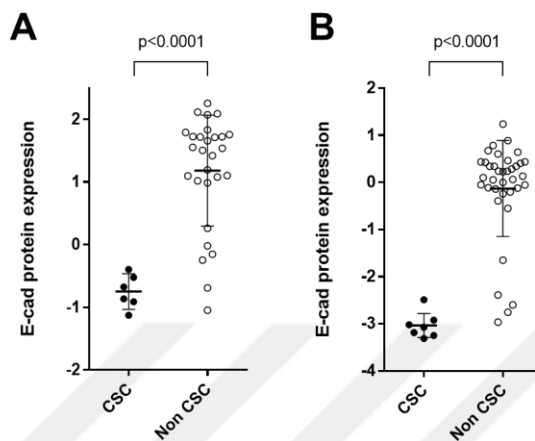


**Figure 5.1: E cadherin (CDH1) expression difference between CSC and non-CSC like clusters.** CDH1 protein expression was significantly higher in non-CSC like phenotypes in TCPA (A) and Marcotte et al. (GSE73526) datasets. Table retrieved from (Akbar et al., 2020a).

In conclusion, CNCL robust classification of breast cancer cell lines into CS/M and NS/E phenotype shows that CS/M group is enriched for mesenchymal related genesets and NS/E group was enriched for epithelial related genesets.

### 5.1.3. CNCL can identify stemness *in silico*

Next, we wanted to evaluate if CNCL could cluster breast cancer based on their stemness. For this purpose, we analyzed CD44, CD24 and ALDH expression in CS/M and NS/E groups identified by CNCL in CCLE and CGP datasets. CD44 and CD24 were differentially expressed between these groups while ALDH did not show any significant difference (Table 5.5).

**Table 5.5: Differential expression of CD44, CD24 and ALDH in CS/M and NS/E groups** Table retrieved from (Akbar et al., 2020a).

| Gene Symbol | CCLE             |             | CGP              |             |
|-------------|------------------|-------------|------------------|-------------|
|             | p value (t-test) | Fold change | p value (t-test) | Fold change |
| CD44        | <0.0001          | 2.62        | 0.005            | 1.89        |
| CD24        | <0.0001          | -4.82       | 0.003            | -3.9        |
| ALDH1       | 0.792            | 0.19        | 0.526            | 0.68        |

As CNCL is composed of 15 genes, the correlation between these genes and stem cell markers (CD44, CD24 and ALDH) was calculated individually. Almost all correlations between CNCL genes and stem cell markers (CD44 and CD24) showed an expected and significant correlation but the same pattern could not be found with ALDH as shown in Table 5.6.

**Table 5.6: CNCL genes showed a significant correlation pattern with CD44, CD24 and ALDH in CCLE and CGP datasets.** Table retrieved from (Akbar et al., 2020a)

| CCLE         | Pearson Correlation |       |       | p value |      |       |
|--------------|---------------------|-------|-------|---------|------|-------|
| Gene symbols | CD44                | CD24  | ALDH1 | CD44    | CD24 | ALDH1 |
| FN1          | 0.57                | -0.73 | -0.07 | 0       | 0    | 0.61  |
| VIM          | 0.69                | -0.67 | -0.04 | 0       | 0    | 0.8   |
| DDR2         | 0.49                | -0.86 | -0.06 | 0       | 0    | 0.69  |
| SLIT2        | 0.47                | -0.66 | 0.08  | 0       | 0    | 0.54  |
| PVRL3        | 0.5                 | -0.74 | 0.07  | 0       | 0    | 0.59  |
| BNC2         | 0.55                | -0.72 | -0.03 | 0       | 0    | 0.82  |
| TMEM158      | 0.61                | -0.74 | -0.14 | 0       | 0    | 0.3   |
| DKK3         | 0.45                | -0.84 | -0.15 | 0       | 0    | 0.28  |
| CLDN4        | -0.47               | 0.73  | -0.06 | 0       | 0    | 0.64  |
| CDH1         | -0.37               | 0.78  | 0.1   | 0       | 0    | 0.48  |
| ST14         | -0.45               | 0.7   | 0.06  | 0       | 0    | 0.68  |
| IRF6         | -0.35               | 0.7   | 0.04  | 0.01    | 0    | 0.79  |
| AP1M2        | -0.43               | 0.73  | -0.01 | 0       | 0    | 0.94  |
| ZNF165       | -0.27               | 0.73  | 0.08  | 0.04    | 0    | 0.54  |
| BSPRY        | -0.53               | 0.78  | -0.12 | 0       | 0    | 0.38  |

| CGP          | Pearson Correlation |       |       | p value |      |       |
|--------------|---------------------|-------|-------|---------|------|-------|
| Gene symbols | CD44                | CD24  | ALDH1 | CD44    | CD24 | ALDH1 |
| FN1          | 0.5                 | -0.57 | -0.06 | 0       | 0    | 0.68  |
| VIM          | 0.6                 | -0.56 | 0.04  | 0       | 0    | 0.81  |
| DDR2         | 0.38                | -0.54 | 0.15  | 0.01    | 0    | 0.35  |
| SLIT2        | 0.21                | -0.38 | 0.33  | 0.17    | 0.01 | 0.03  |
| PVRL3        | 0.35                | -0.66 | 0.31  | 0.02    | 0    | 0.05  |
| BNC2         | 0.34                | -0.58 | -0.01 | 0.03    | 0    | 0.97  |
| TMEM158      | 0.44                | -0.71 | -0.05 | 0       | 0    | 0.73  |
| DKK3         | 0.37                | -0.55 | -0.11 | 0.02    | 0    | 0.51  |
| CLDN4        | -0.08               | 0.48  | -0.08 | 0.6     | 0    | 0.61  |
| CDH1         | -0.32               | 0.69  | 0.19  | 0.04    | 0    | 0.23  |
| ST14         | -0.33               | 0.41  | -0.05 | 0.03    | 0.01 | 0.74  |
| IRF6         | -0.23               | 0.57  | -0.11 | 0.14    | 0    | 0.48  |
| AP1M2        | -0.25               | 0.59  | -0.08 | 0.11    | 0    | 0.6   |
| ZNF165       | -0.23               | 0.33  | -0.13 | 0.15    | 0.03 | 0.41  |
| BSPRY        | -0.13               | 0.63  | -0.22 | 0.41    | 0    | 0.17  |

This analysis clearly showed that a classification based on CD44 and CD24 identifies the same groups as CNCL but ALDH1 expression does not reflect these groups. To identify ALDH relation with CD44 and CD24, we performed Pearson correlation but could not find any significant correlation between these in CCLE or CGP datasets as shown in Table 5.7, which shows that CD44 and CD24 expression are not related to that of ALDH1.

**Table 5.7: CD44 and CD24 correlation with ALDH in CCLE and CGP datasets.** Table retrieved from (Akbar et al., 2020a).

| Pearson Correlation with ALDH | CCLE      |         | CGP       |         |
|-------------------------------|-----------|---------|-----------|---------|
| Gene Symbol                   | Pearson r | p value | Pearson r | p value |
| CD44                          | 0.18      | 0.18    | 0         | 0.99    |
| CD24                          | 0.15      | 0.27    | 0.08      | 0.6     |

As we could show that CNCL can cluster CD44+/CD24- stem cells and CD44-/CD24+ non-stem cells separately from each other which were derived from MCF10A (GSE15192) (Bhat-Nakshatri et al., 2010; Akbar, 2014), we further evaluated the correlation between CD44, CD24 and ALDH1 with CNCL genes in this dataset (GSE15192). This analysis showed the same results as with our previous correlation analysis in CCLE and CGP datasets, CNCL genes were significantly correlated with CD44 and CD24 but not with ALDH1 Table 5.8.

**Table 5.8: CNCL genes correlation with CD44, CD24 and ALDH1. Table retrieved from (Akbar et al., 2020a).**

| GSE15192     | Pearson Correlation |      |       | p value |        |       |
|--------------|---------------------|------|-------|---------|--------|-------|
| Gene symbols | CD44                | CD24 | ALDH1 | CD44    | CD24   | ALDH1 |
| FN1          | 0.72                | -1   | -0.65 | 0.04    | <0.001 | 0.08  |
| VIM          | 0.69                | -1   | -0.61 | 0.06    | <0.001 | 0.11  |
| DDR2         | -0.2                | 0.38 | -0.03 | 0.59    | 0.36   | 0.94  |
| SLIT2        | 0.67                | -1   | -0.72 | 0.07    | <0.001 | 0.04  |
| PVRL3        | 0.68                | -1   | -0.61 | 0.06    | <0.001 | 0.11  |
| BNC2         | 0.64                | -0.9 | -0.75 | 0.09    | <0.001 | 0.03  |
| TMEM158      | 0.71                | -1   | -0.54 | 0.05    | <0.001 | 0.17  |
| DKK3         | 0.36                | -0.8 | -0.43 | 0.38    | 0.03   | 0.29  |
| CLDN4        | -0.8                | 1    | 0.61  | 0.03    | <0.001 | 0.11  |
| CDH1         | -0.7                | 1    | 0.62  | 0.06    | <0.001 | 0.1   |
| ST14         | -0.7                | 1    | 0.59  | 0.05    | <0.001 | 0.13  |
| IRF6         | -0.7                | 1    | 0.61  | 0.07    | <0.001 | 0.11  |
| AP1M2        | -0.7                | 0.99 | 0.62  | 0.08    | <0.001 | 0.1   |
| ZNF165       | -0.7                | 0.99 | 0.59  | 0.08    | <0.001 | 0.13  |
| BSPRY        | -0.6                | 0.99 | 0.62  | 0.1     | <0.001 | 0.1   |

Further, we showed that CNCL could differentiate between stem and non-stem cells (GSE36643) (Battula et al., 2012) which were isolated from HMLER cell line by FACS using CD44 and CD24 markers as shown in Figure 5.2.

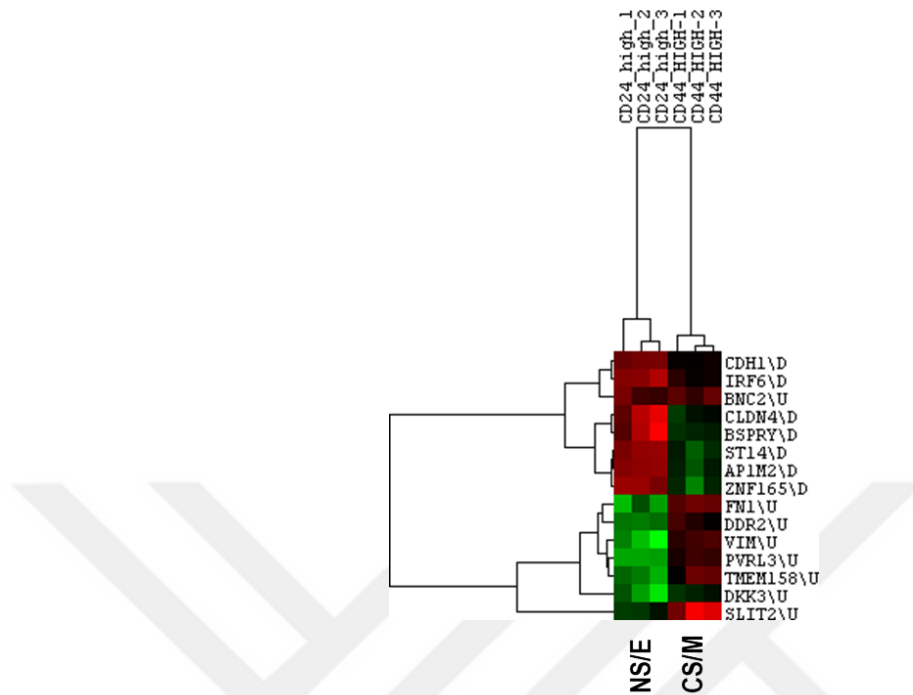


**Figure 5.2: CNCL clustering in dataset GSE36643.** CNCL classified CD44 high cells as CS/M and CD24 high cells as NS/E from HMLER cells isolated by FACS sorting as expected. Figure retrieved from (Akbar et al., 2020a).

In another analysis, CNCL genes were highly correlated with CD44 and CD24 as expected but not with ALDH1 in GSE36643 upon Pearson correlation analysis as shown in Table 5.9.

**Table 5.9: Pearson correlation analysis between CNCL genes and stemness genes (CD44 and CD24) Table retrieved from (Akbar et al., 2020a).**

| Gene symbols | Pearson Correlation |       |       | p value |        |       |
|--------------|---------------------|-------|-------|---------|--------|-------|
|              | CD44                | CD24  | ALDH1 | CD44    | CD24   | ALDH1 |
| FN1          | -0.91               | -0.88 | -0.77 | 0.01    | 0.02   | 0.07  |
| VIM          | -0.9                | -0.99 | -0.69 | 0.02    | <0.001 | 0.13  |
| DDR2         | -0.99               | -0.94 | -0.93 | <0.001  | 0.01   | 0.01  |
| SLIT2        | -0.88               | -0.85 | -0.73 | 0.02    | 0.03   | 0.1   |
| PVRL3        | -0.95               | -0.94 | -0.81 | <0.001  | 0.01   | 0.05  |
| BNC2         | 0.09                | -0.25 | 0.25  | 0.87    | 0.63   | 0.63  |
| TMEM158      | -0.85               | -0.92 | -0.62 | 0.03    | 0.01   | 0.19  |
| DKK3         | -0.74               | -0.93 | -0.47 | 0.1     | 0.01   | 0.34  |
| CLDN4        | 0.9                 | 0.99  | 0.73  | 0.02    | <0.001 | 0.1   |
| CDH1         | 0.97                | 0.98  | 0.84  | <0.001  | <0.001 | 0.04  |
| ST14         | 0.95                | 0.95  | 0.8   | <0.001  | <0.001 | 0.06  |
| IRF6         | 0.9                 | 0.96  | 0.69  | 0.02    | <0.001 | 0.13  |
| AP1M2        | 0.97                | 0.94  | 0.84  | <0.001  | <0.001 | 0.04  |
| ZNF165       | 0.96                | 0.88  | 0.86  | <0.001  | 0.02   | 0.03  |
| BSPRY        | 0.84                | 0.97  | 0.63  | 0.03    | <0.001 | 0.18  |

Next, we selected a dataset (GSE52327) (Liu et al., 2014b) in which cells were selected for ALDH1 expression using flow cytometry. CNCL gene list separation of ALDH1+ cells from ALDH1- was incomplete, supporting our previous results that CD44 and CD24 identify different cell state as compared to ALDH1 (Figure 5.3).

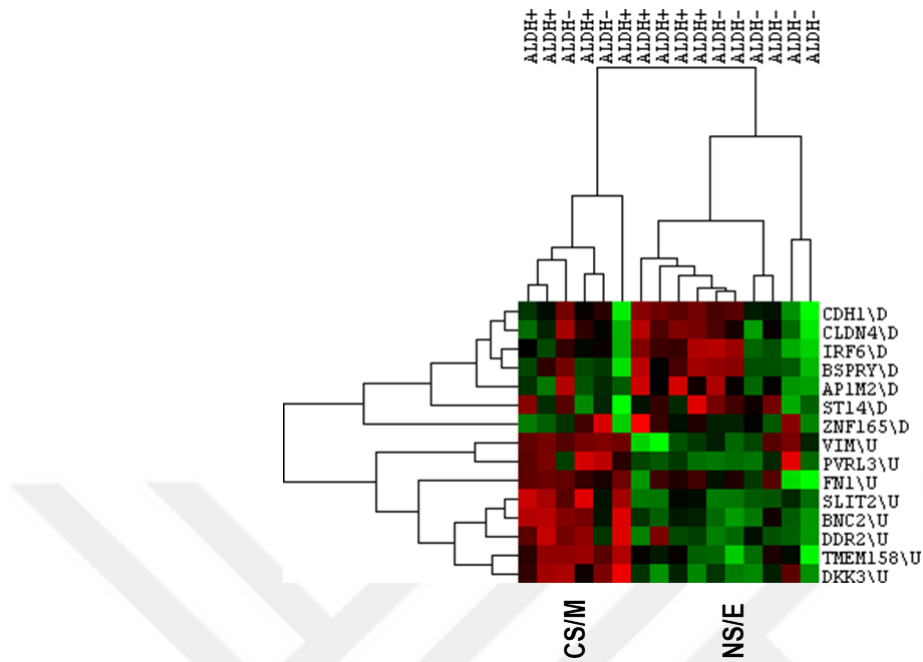


**Figure 5.3. Clustering of primary breast cancer cells sorted by ALDH1 expression with CNCL.** CNCL based separation could not show a clear separation between ALDH+ and ALDH- cells sorted by flow cytometry (GSE52327). Figure retrieved from (Akbar et al., 2020a).

These results show that CNCL gene list can classify cells into CS/M (stem cells) and NS/E (non-stem cells) subgroups when correlated with CD44 and CD24 markers, but not if ALDH1 is used as a marker.

#### 5.1.4. CNCL can identify EMT *in silico*

As CNCL genes contain genes that have been closely related to EMT we asked if CNCL could differentiate between epithelial and mesenchymal cell states. In dataset GSE9691 (Onder et al., 2008) HMLER cells were treated with shRNA to knockdown E-cad alone, or both  $\beta$ -Cat (beta catenin) and E-cad, or cells were transfected with a dominant negative version of E-cad. CNCL classified cells, in which E-Cad was downregulated using shRNA or transfection, as CS/M, and cells in which E-cad downregulation was rescued with  $\beta$ -Cat downregulation along with the control cells as NS/E as expected (Figure 5.4). E-cad expression was significantly different between cells representing mesenchymal state CS/M and epithelial state NS/E.

Increased E-cad is known to cause MET. This has been related to its blockage of  $\beta$ -catenin ( $\beta$ -Cat) function. Therefore, when E-cad is downregulated,  $\beta$ -Cat is activated causing EMT. This can be reversed if  $\beta$ -Cat is also downregulated (Schmalhofer et al., 2009) (Kim et al., 2002) (Yang et al., 2006) (Eger et al., 2004) (Morali et al., 2001)

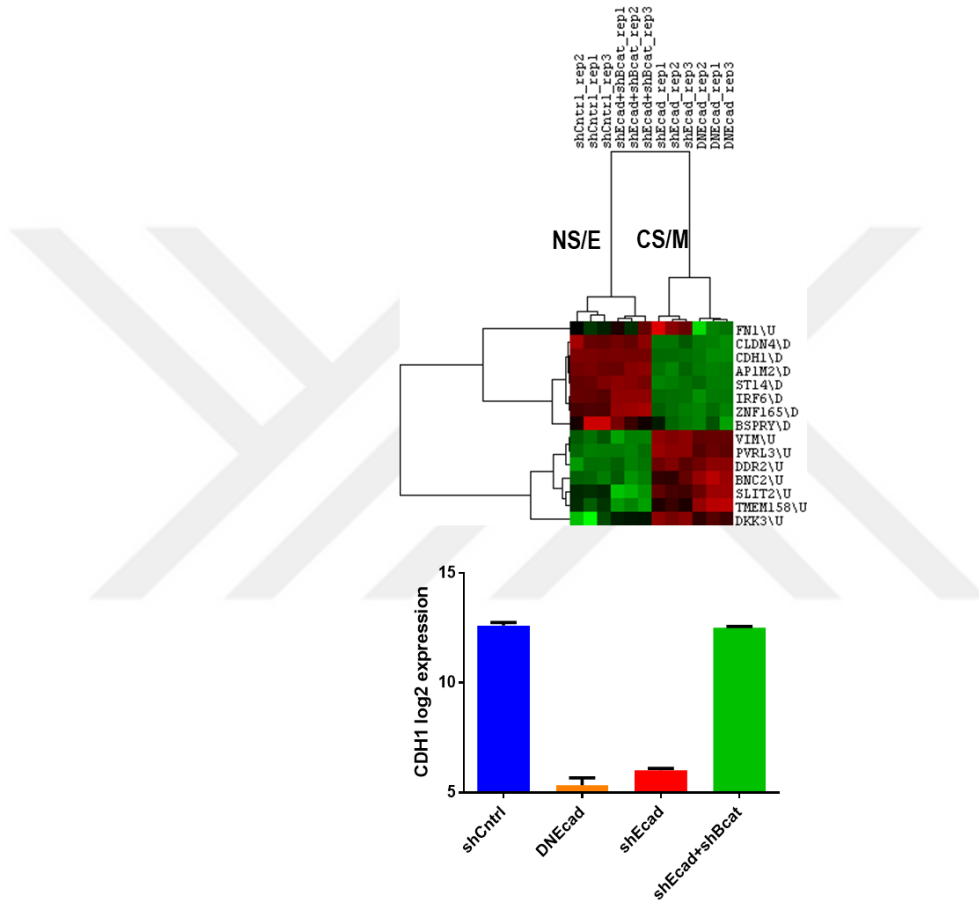


**Figure 5.4: CNCL clustering in dataset GSE9691.** CNCL could classify mesenchymal phenotype cells which were induced by either E-cad downregulation or dominant negative E-cad version transfection from epithelial phenotype represented by control cells and cells where both E-cad and  $\beta$ -Cat were downregulated. The difference between both states was significantly different as shown in the bar graph. Figure retrieved from (Akbar et al., 2020a).

In another dataset (GSE24202) (Taube et al., 2010), HMLER cells were identified as CS/M phenotype when these were retrovirally transfected with TGF $\beta$ , Twist, GSC or Snail, as expected by CNCL clustering (Figure 5.5). Unexpectedly cells in which E-cad was downregulated using shRNA, were identified as NS/E subtype. However, this is most likely

due to the incomplete down-regulation resulting from the shRNA used in this study, and EMT requires a more efficient down-regulation of E-cad.

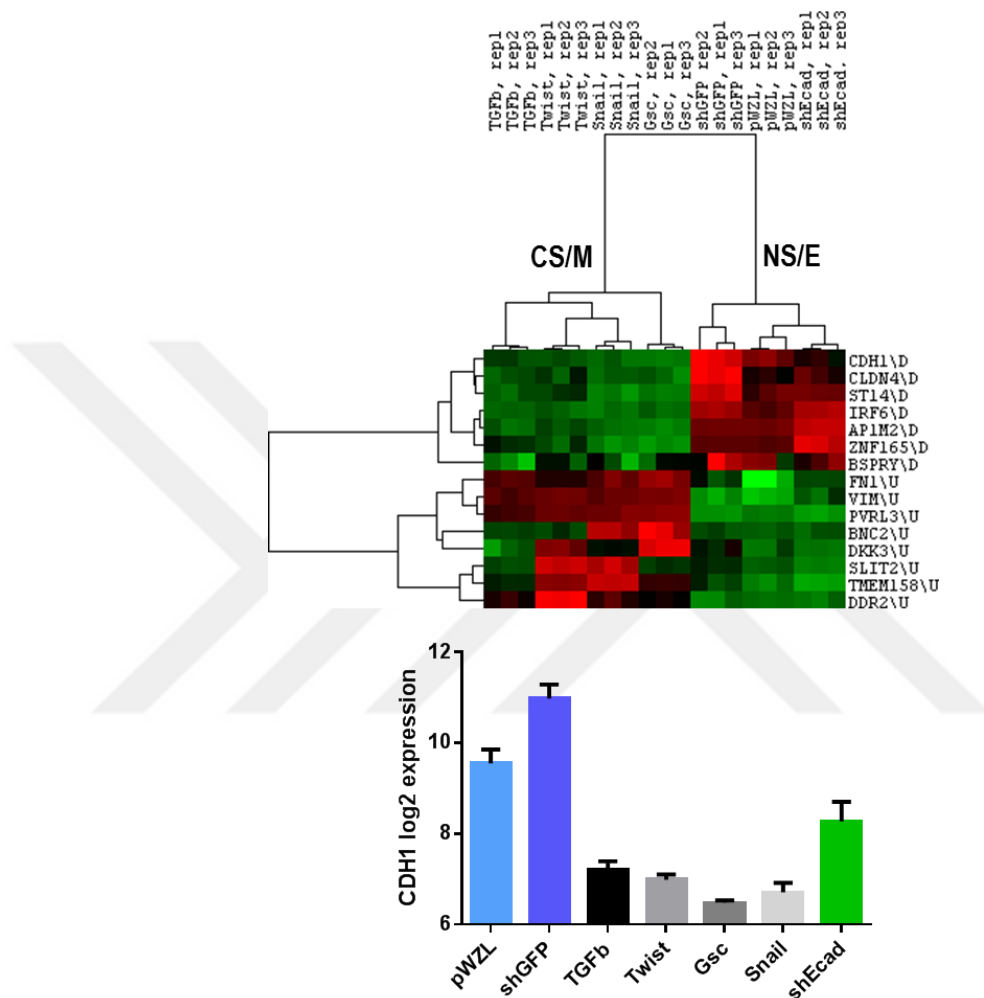


**Figure 5.5: CNCL clustering in dataset GSE24202.** CNCL classification showed cells overexpressing TGF $\beta$ , Twist, GSC or Snail as CS/M when compared to controls. Figure retrieved from (Akbar et al., 2020a).

In another dataset GSE7515 (Creighton et al., 2009), the authors generated mammospheres from primary breast cancer cells. CNCL classified 12/15 mammospheres as CS/M phenotype as expected and 10/11 primary cancer cell lines were clustered as NS/E and intermediary phenotype as expected (Figure 5.6). Exceptions to CNCL classification can be attributed to the fact that some primary cell lines can be mesenchymal to start with and mammosphere formation does not always lead to strong induction of EMT.

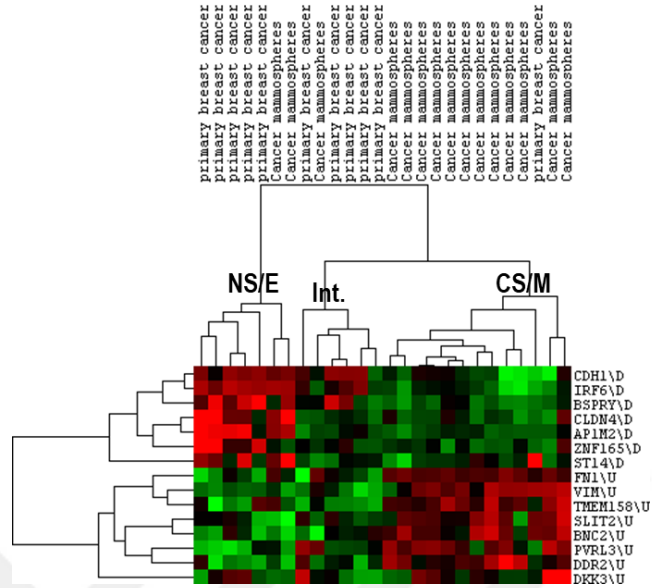


**Figure 5.6: CNCL clustering in dataset GSE7515.** CNCL classified most mammospheres (12/15) as CS/M and primary cell lines (10/11) as either NS/E or intermediary as expected. Figure retrieved from (Akbar et al., 2020a).

These results suggest CNCL can identify the mesenchymal state as CS/M and epithelial state of cells as NS/E phenotype in both cell lines and primary cancer cells.

#### 5.1.5. CNCL can predict drug sensitivity *in silico*

In literature, both CSC and mesenchymal cancer cells have been reported as having high resistance to the majority of chemo and targeted therapy regimens. To test if CNCL can cluster sensitive cells from resistant cells, further analysis was performed next. In GSE24460 (Calcagno et al., 2010), MCF7 cells were made resistant to doxorubicin with prolonged treatment. CNCL classified resistant MCF7 cells as CS/M and control/untreated cells were clustered as NS/E as expected (Figure 5.7).

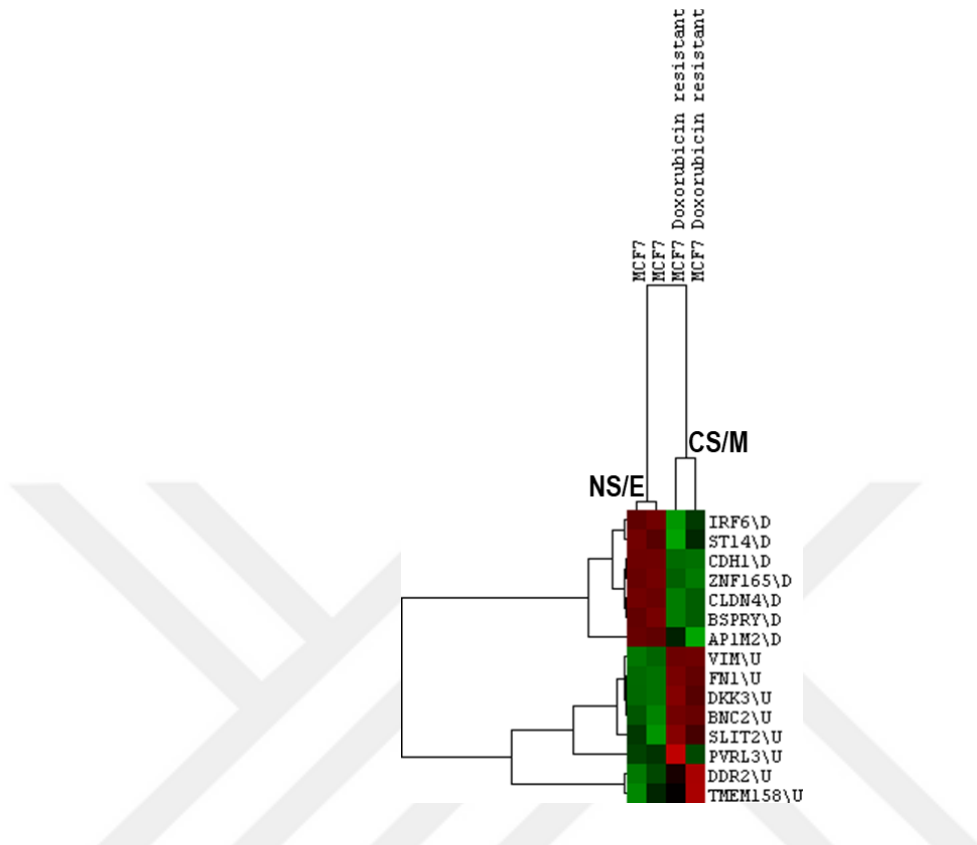


**Figure 5.7: CNCL clustering in GSE24460.** CNCL classified MCF7 doxorubicin resistant cells as CS/M and MCF7 control cells as NS/E as expected. Figure retrieved from (Akbar et al., 2020a).

Additionally, breast cancer tumors develop resistance upon treatment with different therapies. In dataset GSE10281 (Creighton et al., 2009), 18 patients were treated for 3 months with letrozole and their pre- and post-treatment expression data was available. Analysis with CNCL showed that most of the pretreatment samples showed NS/E phenotype and among post treatment samples, many showed CS/M phenotype. Interestingly when matched samples from each patient were compared, 7 tumor samples which were classified as NS/E transitioned to CS/M state and only 1 pretreatment sample with CS/M phenotype transitioned to NS/E phenotype (Figure 5.8 and Table 5.10).

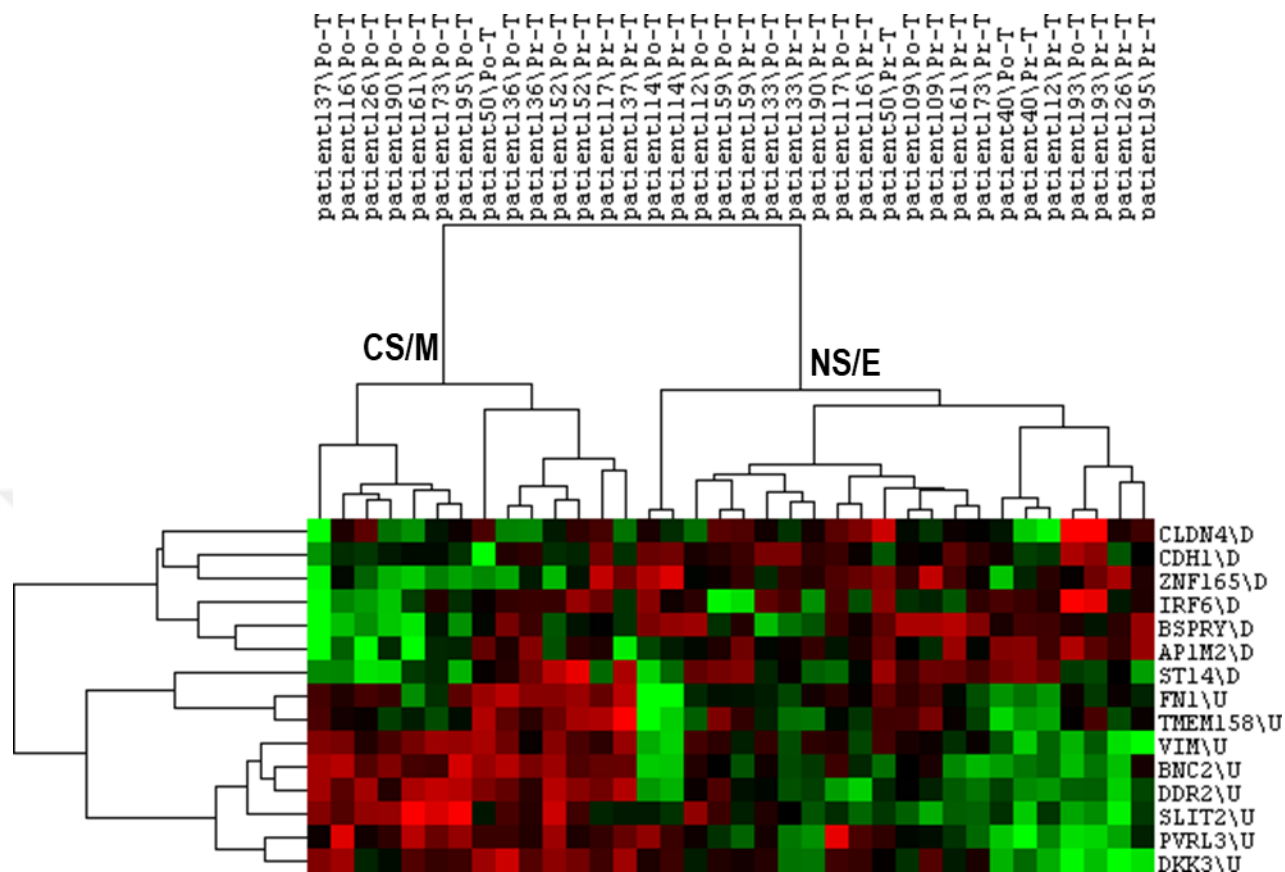


**Figure 5.8: CNCL clustering in GSE10281.** CNCL clustered the majority of pre letrozole treated samples as NS/E and post treated samples as CS/M. Figure retrieved from (Akbar et al., 2020a).

**Table 5.10: Phenotype comparison of matched tumor samples pre and post treatment with letrozole.** Table retrieved from (Akbar et al., 2020a)

| Patient ID | Pre-TRX | Post-TRX |
|------------|---------|----------|
| 50         | NS/E    | CS/M     |
| 116        | NS/E    | CS/M     |
| 126        | NS/E    | CS/M     |
| 161        | NS/E    | CS/M     |
| 173        | NS/E    | CS/M     |
| 190        | NS/E    | CS/M     |
| 195        | NS/E    | CS/M     |
| 40         | NS/E    | NS/E     |
| 109        | NS/E    | NS/E     |
| 112        | NS/E    | NS/E     |
| 114        | NS/E    | NS/E     |
| 133        | NS/E    | NS/E     |
| 159        | NS/E    | NS/E     |
| 193        | NS/E    | NS/E     |
| 136        | CS/M    | CS/M     |
| 137        | CS/M    | CS/M     |
| 152        | CS/M    | CS/M     |
| 117        | CS/M    | NS/E     |

These results show that tumors with NS/E phenotype can transition to CS/E phenotype resulting in resistance to treatment.

#### 5.1.6. Stemness score (SS) based quantitative analyses

To trace the change in CS/M and NSC/E phenotype we assigned a quantitative value of stemness to each sample in a dataset, by generating a CNCL stemness matrix. The matrix consisted of 2 columns, CS/M and NS/E, with each containing reference values corresponding to the median values obtained for each gene in the CNCL list, derived from clustered cell lines in that respective group after mean standardization in the CCLE dataset (Table 5.11). For every sample, the expression of all genes within the CNCL gene list was

Pearson correlated with the reference CS/M and NS/E values generating CS/M(r) and NS/E(r) scores.

**Table 5.11: CNCL stemness reference matrix** Table retrieved from (Akbar et al., 2020a)

(\*shows mean standardized expression values in CS/M cells and \*\* shows mean standardized expression values in NS/E cells in CCLE dataset)

| Gene symbol/Probeset ID | Gene related to | CS/M matrix* | NS/E matrix** |
|-------------------------|-----------------|--------------|---------------|
| DDR2\U\205168_at        | CS/M            | 1.87         | -0.507        |
| DKK3\U\202196_s_at      | CS/M            | 1.866        | -0.524        |
| SLIT2\U\209897_s_at     | CS/M            | 1.698        | -0.588        |
| PVRL3\U\213325_at       | CS/M            | 1.601        | -0.7          |
| BNC2\U\220272_at        | CS/M            | 1.544        | -0.662        |
| TMEM158\U\213338_at     | CS/M            | 1.534        | -0.689        |
| FN1\U\212464_s_at       | CS/M            | 1.498        | -0.312        |
| VIM\U\201426_s_at       | CS/M            | 1.351        | -0.815        |
| ST14\D\202005_at        | NS/E            | -1.447       | 0.544         |
| IRF6\D\202597_at        | NS/E            | -1.496       | 0.509         |
| BSPRY\D\218792_s_at     | NS/E            | -1.539       | 0.625         |
| ZNF165\D\206683_at      | NS/E            | -1.566       | 0.537         |
| CLDN4\D\201428_at       | NS/E            | -1.572       | 0.515         |
| CDH1\D\201131_s_at      | NS/E            | -1.672       | 0.717         |
| AP1M2\D\65517_at        | NS/E            | -1.807       | 0.518         |

This approach classified paclitaxel sensitive (both control and paclitaxel resensitized cells after bexarotene treatment) and resistant MDA-MB-231 cells in dataset GSE12791 (Luo et al., 2011) as shown in Figure 5.9.

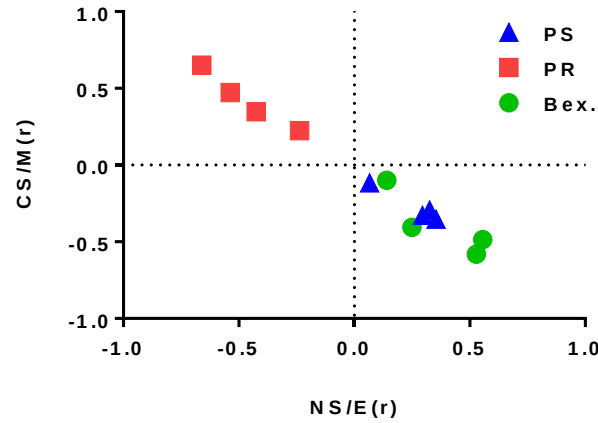


**Figure 5.9: Stemness score-based clustering in dataset GSE12791.** MDA-MB-231 cells which were grown resistant to paclitaxel over a prolonged treatment, showed higher CS/M(r) and lower NS/E(r) when compared to sensitive cells (control cells/ cells which were re-sensitized to paclitaxel after bexarotene treatment). Figure retrieved from (Akbar et al., 2020a).

A similar phenotypic classification was observed when another dataset (GSE23399) (Chan et al., 2016) was analyzed containing expression data of cancer-associated fibroblast (CAF) treated with paclitaxel for 1, 3 or 7 days. CAF showed higher CS/M(r) and lower NS/E(r) with an increase in paclitaxel treatment time (Figure 5.10).

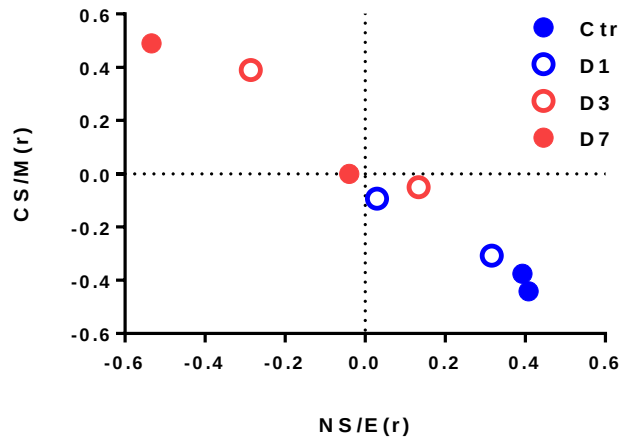


**Figure 5.10: Stemness score-based clustering in dataset GSE23399.** CAFs showed an increase in CS/M(r) and a decrease in NS/E(r) as the cells were treated with paclitaxel for a long period. Figure retrieved from (Akbar et al., 2020a).

In another dataset (GSE 16179) (Liu et al., 2009), BT20 cells were made resistant to Lapatinib and CNCL stemness score based analysis showed that resistant cells exhibited higher CS/M(r) and lower NS/E(r) when compared to sensitive cells as shown in Figure 5.11.

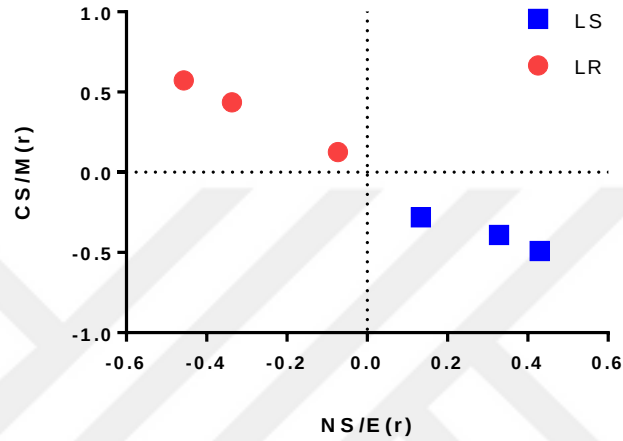


**Figure 5.11: Stemness score-based clustering in dataset GSE16179.** Lapatinib BT20 resistant cells showed increased CS/M(r) and decreased NS/E(r) when compared to sensitive cells. Figure retrieved from (Akbar et al., 2020a).

These results indicate that stemness as evaluated by CS/M(r) and NS/E(r) is indeed a dynamic and reversible process. We next asked if the same could be the case in tumors. For this we generated a single value from CS/M(r) and NS/E(r) by calculating SS Delta(r) as shown below where an increase in Delta(r) showed an increase in stemness/mesenchymalness and vice versa.

$$\text{Delta}(r) = \text{CS/M}(r) - \text{NS/E}(r)$$

GSE28844 (Vera-Ramirez et al., 2013) contained patient data before and after Anthracycline and Taxane based neo-adjuvant chemotherapy for 29 breast cancer patients. Upon Delta(r) based analysis, post-therapy patient samples showed increased Delta(r) when compared to pre-therapy samples, suggesting a general phenotypic switch induced from therapy in all tumors (Figure 5.12). There was no matching information available between pre and post therapy tumors so the phenotypic transition in each patient can't be addressed in this dataset.

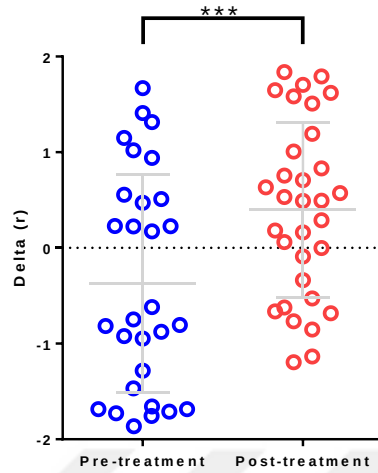


**Figure 5.12: Delta(r) based analysis in GSE28844.** Patient tumors post Anthracycline and Taxane based neo-adjuvant showed higher Delta (r) when compared with pre-therapy tumor samples. Figure retrieved from (Akbar et al., 2020a).

These analyses show that similar to our previous CNCL clustering approaches, both the stemness score and the Delta(r) approach can identify a dynamic change between samples in which resistance is induced from prolonged exposure to drugs when compared with control or pre-therapy samples.

#### 5.1.7. CNCL validation *in vitro*

To validate CNCL based classification *in vitro*, 18 different breast cancer cell lines were used to determine CNCL gene expression values using qPCR after which we performed the below described experiments. Indeed, gene expression data obtained through qPCR was highly concordant with the CCLE gene expression data (Table 5.12). Additionally, selected genes showed the expected pattern of intergenic correlations (Table 5.13). Using qPCR data, we generated a SS matrix and ultimately Delta(r) values for further experiments.

**Table 5.12: qPCR data was highly concordant with CCLE gene expression data.**

Table retrieved from (Akbar et al., 2020a)

| Gene    | Pearson<br>r | p value |
|---------|--------------|---------|
| VIM     | 0.94         | <0.0001 |
| ST14    | 0.93         | <0.0001 |
| CDH1    | 0.93         | <0.0001 |
| AP1M2   | 0.92         | <0.0001 |
| IRF6    | 0.91         | <0.0001 |
| BSPRY   | 0.89         | <0.0001 |
| DKK3    | 0.81         | <0.0001 |
| PVRL3   | 0.8          | <0.0001 |
| BNC2    | 0.78         | 0.0002  |
| FN1     | 0.72         | <0.001  |
| TMEM158 | 0.7          | 0.002   |
| DDR2    | 0.69         | 0.002   |
| SLIT2   | 0.69         | 0.002   |
| ZNF165  | 0.64         | <0.01   |
| CLDN4   | 0.46         | 0.06    |

**Table 5.13: CNCL genes showed a strong intergenic correlation pattern as expected.**

Table retrieved from (Akbar et al., 2020a)

| qPCR    | FN1    | VIM    | DDR2   | SLIT2  | PVRL3  | BNC2   | TMEM158 | DKK3   | CLDN4  | CDH1   | ST14   | IRF6   | AP1M2  | ZNF165 | BSPRY  |
|---------|--------|--------|--------|--------|--------|--------|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| FN1     | 1.000  | 0.589  | 0.667  | 0.278  | 0.264  | 0.524  | 0.135   | 0.683  | 0.125  | -0.232 | -0.296 | -0.091 | -0.560 | 0.186  | -0.175 |
| VIM     | 0.589  | 1.000  | 0.701  | 0.552  | 0.655  | 0.519  | 0.533   | 0.591  | -0.126 | -0.587 | -0.573 | -0.405 | -0.534 | -0.027 | -0.481 |
| DDR2    | 0.667  | 0.701  | 1.000  | 0.556  | 0.723  | 0.492  | 0.184   | 0.690  | -0.055 | -0.526 | -0.583 | -0.459 | -0.699 | -0.109 | -0.511 |
| SLIT2   | 0.278  | 0.552  | 0.556  | 1.000  | 0.729  | 0.385  | 0.603   | 0.589  | 0.028  | -0.458 | -0.449 | -0.402 | -0.390 | -0.102 | -0.522 |
| PVRL3   | 0.264  | 0.655  | 0.723  | 0.729  | 1.000  | 0.181  | 0.489   | 0.446  | -0.210 | -0.644 | -0.678 | -0.591 | -0.534 | -0.287 | -0.649 |
| BNC2    | 0.524  | 0.519  | 0.492  | 0.385  | 0.181  | 1.000  | 0.271   | 0.608  | 0.015  | -0.314 | -0.416 | -0.270 | -0.449 | 0.116  | -0.316 |
| TMEM158 | 0.135  | 0.533  | 0.184  | 0.603  | 0.489  | 0.271  | 1.000   | 0.405  | 0.159  | -0.422 | -0.307 | -0.157 | -0.009 | 0.121  | -0.273 |
| DKK3    | 0.683  | 0.591  | 0.690  | 0.589  | 0.446  | 0.608  | 0.405   | 1.000  | 0.404  | -0.151 | -0.199 | 0.020  | -0.470 | 0.466  | -0.124 |
| CLDN4   | 0.125  | -0.126 | -0.055 | 0.028  | -0.210 | 0.015  | 0.159   | 0.404  | 1.000  | 0.609  | 0.575  | 0.668  | 0.514  | 0.798  | 0.619  |
| CDH1    | -0.232 | -0.587 | -0.526 | -0.458 | -0.644 | -0.314 | -0.422  | -0.151 | 0.609  | 1.000  | 0.937  | 0.903  | 0.591  | 0.545  | 0.934  |
| ST14    | -0.296 | -0.573 | -0.583 | -0.449 | -0.678 | -0.416 | -0.307  | -0.199 | 0.575  | 0.937  | 1.000  | 0.932  | 0.680  | 0.526  | 0.901  |
| IRF6    | -0.091 | -0.405 | -0.459 | -0.402 | -0.591 | -0.270 | -0.157  | 0.020  | 0.668  | 0.903  | 0.932  | 1.000  | 0.566  | 0.688  | 0.939  |
| AP1M2   | -0.560 | -0.534 | -0.699 | -0.390 | -0.534 | -0.449 | -0.009  | -0.470 | 0.514  | 0.591  | 0.680  | 0.566  | 1.000  | 0.345  | 0.614  |
| ZNF165  | 0.186  | -0.027 | -0.109 | -0.102 | -0.287 | 0.116  | 0.121   | 0.466  | 0.798  | 0.545  | 0.526  | 0.688  | 0.345  | 1.000  | 0.644  |
| BSPRY   | -0.175 | -0.481 | -0.511 | -0.522 | -0.649 | -0.316 | -0.273  | -0.124 | 0.619  | 0.934  | 0.901  | 0.939  | 0.614  | 0.644  | 1.000  |

Previously Midostaurin was found to be preferentially cytotoxic for CS/M cell lines and Lapatinib was cytotoxic for NS/E cell lines *in silico* (Akbar, 2014). This observation was here validated in 18 different cell lines using CyQuant (Figure 5.13 A) and CellTiter Glo assays (Figure 5.13 B).

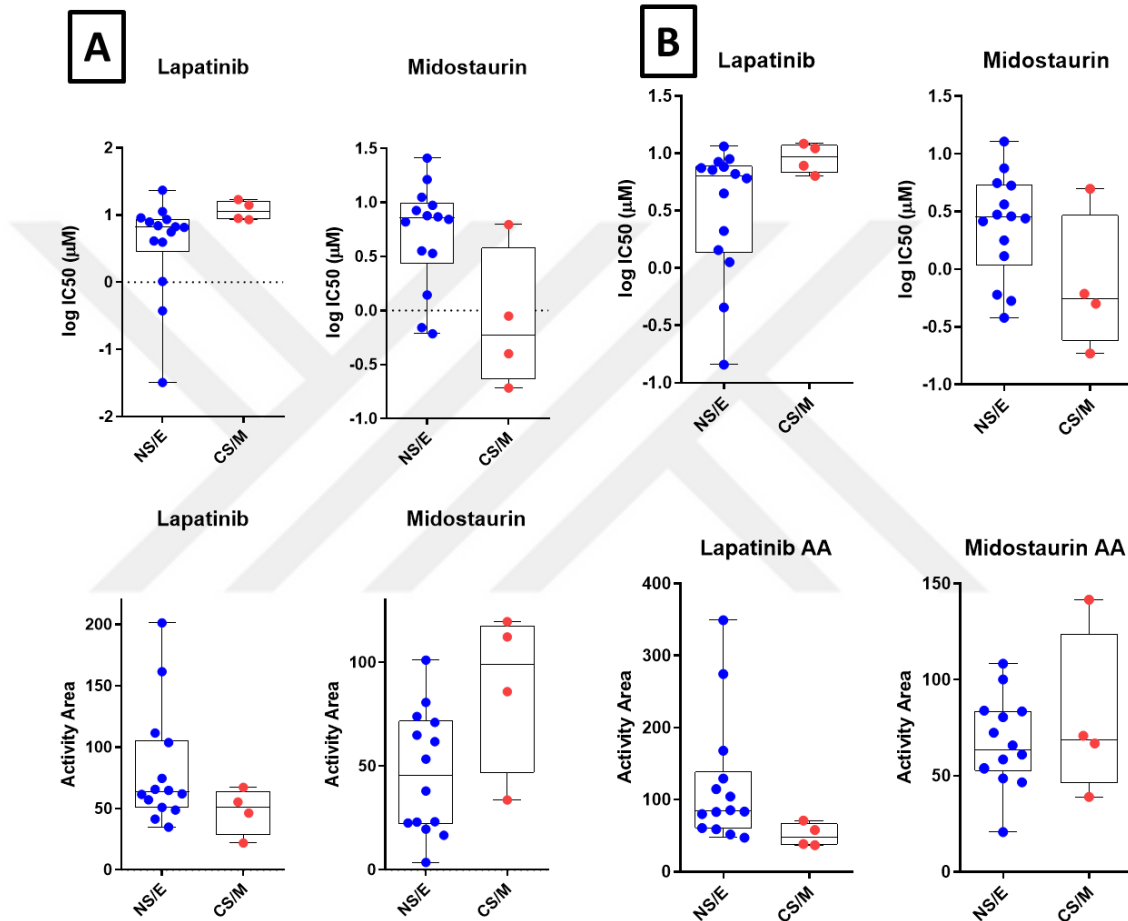


**Figure 5.13: Midostaurin shows preferential cytotoxicity for CS/M cell lines and Lapatinib shows preferential cytotoxicity to NS/E cell lines. (A)** CyQuant assay (t.test p values are Lapatinib IC50: 0.19; Lapatinib AA: 0.19; Midostaurin IC50: 0.01; Midostaurin AA: 0.03) **(B)** CellTiter Glo assay (Lapatinib IC50: 0.15; Lapatinib AA: 0.14; Midostaurin IC50: 0.08; Midostaurin AA: 0.4). Figure retrieved from (Akbar et al., 2020a).

To test if phenotype switching *in vitro* can be detected by CNCL, mammospheres from MDA-MB-157 cell line were generated using 3D cell culture and gene expression of CNCL was measured using qPCR. Using CellTiter Glo assay, cytotoxicity of both Midostaurin

and Lapatinib was measured in both 2D and 3D cell cultures. For the MDA-MB-157 cell line, Delta (r) increased from -0.145 (2D cell culture) to +0.145 (3D cell culture) reflecting a shift towards CS/M phenotype. Additionally, Lapatinib sensitivity also decreased with this switch and Midostaurin sensitivity increased as expected (Figure 5.14). These results suggest that mammosphere generation from MDA-MB-157 cells causes a CNCL quantifiable phenotypic change which in turn causes a change in drug sensitivity.

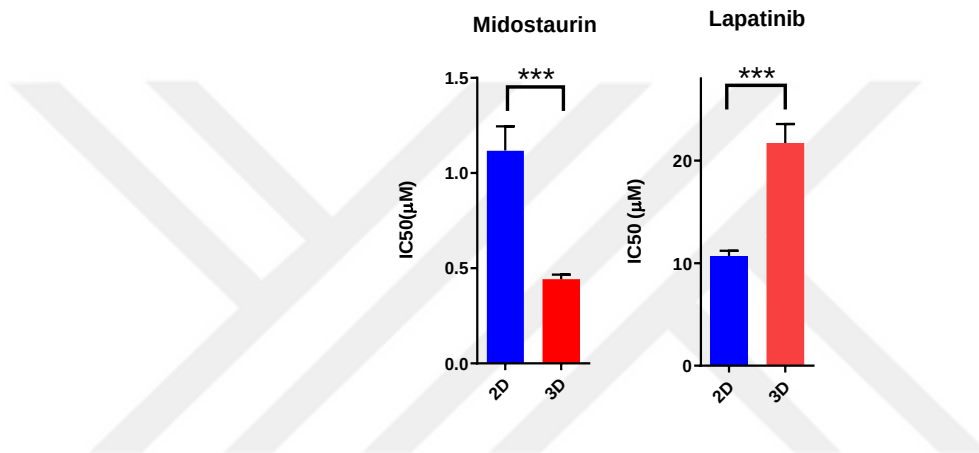


**Figure 5.14: Spheroid cytotoxicity alteration due to phenotypic switch. MDA-MB-157 showed increased sensitivity to Midostaurin and resistance to Lapatinib upon spheroid generation.** Figure retrieved from (Akbar et al., 2020a).

Upon ZEB1 and SNAI2 knockdown CSC like/mesenchymal cells switch to non-CSC like/epithelial phenotype. So, these genes were downregulated in MDA-MB-157 cells using shRNAs (Figure 5.15). Upon gene downregulation, cells were checked for CNCL gene expression and cytotoxicity change for Lapatinib and Midostaurin. Delta(r) changed from +5.1 to -0.12 and -0.01 upon ZEB1 and SNAI2 knockdown respectively depicting a switch from NS/E to CS/M as expected. This phenotype switch also showed increased sensitivity to Lapatinib and resistance to Midostaurin as expected (Figure 5.16). The increase in Lapatinib sensitivity was higher for ZEB1 knockdowns when compared to SNAI2, depicting stronger phenotypic shift which was observed by lower Delta (r).

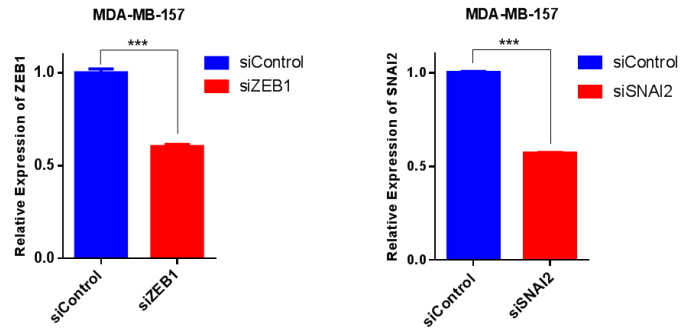


**Figure 5.15: SiZEB1 and SiSNAI2 knockdown efficiency in MDA-MB-157 cell line.** Figure retrieved from (Akbar et al., 2020a).

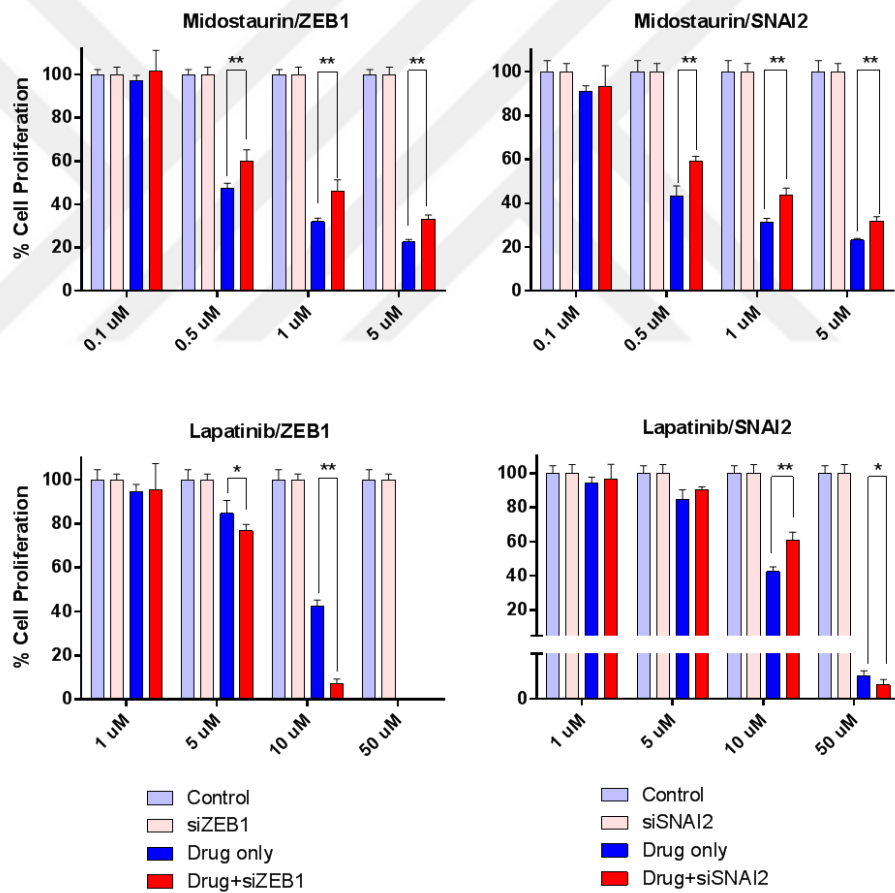


**Figure 5.16: ZEB1 and SNAI2 knockdown resulted in increased sensitivity to Lapatinib and resistance to Midostaurin in MDA-MB-157 cell line as expected.** Figure retrieved from (Akbar et al., 2020a).

Cumulatively these results show that CS/M cells are sensitive to Midostaurin and resistance to Lapatinib when compared to NS/E cells and if these phenotypes are switched then this phenotypic change results in drug sensitivity change as well.

#### **5.1.8. Prognostic significance of CNCL**

To investigate the prognostic importance of CNCL in tumor samples, we calculated Delta (r) values for each sample within multiple cohorts: Metabric cohorts (British (n=994) and Canadian (n=997) (Curtis et al., 2012)), GSE1456 (n=159) (Pawitan et al., 2005), GSE2034 (n=286) (Wang et al., 2005), GSE2603 (n=82) (Minn et al., 2005), GSE3494 (n=251) (Miller et al., 2005), GSE4922 (n=249) (Ivshina et al., 2006), GSE6532 (n=380) (Loi et al., 2007), GSE7390 (n=198) (Desmedt et al., 2007), GSE11121(n=200) (Schmidt et al., 2008), GSE12276 (n=204) (Bos et al., 2009), GSE19615 (n=115) (Li et al., 2010), GSE20685(n=327) (Kao et al., 2011), GSE21653 (n=266) (Sabatier et al., 2011), GSE25066(n=508) (Hatzis et al., 2011) and GSE58812 (n=107) (Jezequel et al., 2015). Delta (r) based analysis of 15 patient cohorts with 29 different end point events, showed significant difference between CS/M and NS/E groups in only 7 tests (Table 5.14). In GSE25066 and GSE2603 patients with CS/M phenotype showed worse prognosis when compared to NS/E phenotype.

**Table 5.14: Prognostic evaluation of 16 breast cancer cohorts based upon SS.** Table retrieved from (Akbar et al., 2020a).

| Dataset                  | End point measure | Patient numbers |                 |      |                 | Median survival time (years) |      | Overall Median Survival Years | Hazard ratio | Cox p value  | Summary of treatment protocol      |
|--------------------------|-------------------|-----------------|-----------------|------|-----------------|------------------------------|------|-------------------------------|--------------|--------------|------------------------------------|
|                          |                   | CS/M            | CS/M (censored) | NS/E | NS/E (censored) | CS/M                         | NS/E |                               |              |              |                                    |
| GSE1456                  | DSS               | 129             | 88%             | 30   | 57%             | NA                           | NA   | NA                            | 0.524        | <b>0.004</b> | Adjuvant tamoxifen therapy         |
|                          | OS                | 132             | 81%             | 27   | 44%             | NA                           | NA   | NA                            | 0.557        | <b>0.002</b> |                                    |
|                          | RFS               | 129             | 81%             | 30   | 50%             | NA                           | NA   | NA                            | 0.579        | <b>0.004</b> |                                    |
| GSE2034                  | BR                | 57              | 93%             | 229  | 97%             | NA                           | NA   | NA                            | 1.263        | 0.52         | No neoadjuvant or adjuvant therapy |
|                          | Relapse           | 143             | 59%             | 143  | 66%             | NA                           | NA   | NA                            | 1.073        | 0.53         |                                    |
| GSE2603                  | MFS               | 30              | 50%             | 52   | 77%             | 6.8                          | NA   | NA                            | 1.426        | 0.14         | ND                                 |
|                          | LNMFs             | 23              | 65%             | 59   | 90%             | NA                           | NA   | NA                            | <b>2.017</b> | <b>0.03</b>  |                                    |
|                          | BMFS              | 24              | 71%             | 58   | 88%             | NA                           | NA   | NA                            | 1.569        | 0.17         |                                    |
| GSE3494                  | DSS               | 57              | 86%             | 194  | 68%             | NA                           | NA   | NA                            | 0.802        | 0.08         | Adjuvant chemotherapy              |
| GSE4922                  | OS                | 203             | 68%             | 46   | 48%             | NA                           | NA   | NA                            | 0.808        | 0.059        | Adjuvant tamoxifen therapy         |
| GSE6532                  | DMFS              | 64              | 89%             | 316  | 72%             | 12.6                         | 12.5 | 12.6                          | 0.754        | <b>0.011</b> | ND                                 |
| GSE7390                  | DMFS              | 163             | 67%             | 35   | 74%             | 18.6                         | NA   | 19.7                          | 0.985        | 0.92         | No therapy                         |
|                          | OS                | 58              | 76%             | 140  | 70%             | NA                           | NA   | NA                            | 0.968        | 0.83         |                                    |
|                          | RFS               | 58              | 76%             | 140  | 70%             | 18.1                         | 13.5 | 16                            | 0.968        | 0.83         |                                    |
|                          | TDM               | 55              | 80%             | 143  | 72%             | 19.7                         | NA   | 19.7                          | 0.915        | 0.57         |                                    |
| GSE11121                 | DMFS              | 69              | 88%             | 131  | 71%             | NA                           | NA   | NA                            | 0.763        | 0.08         | Only surgery                       |
| GSE12276                 | RFS               | 21              | 10%             | 183  | 9%              | 1.9                          | 1.8  | 1.8                           | 0.981        | 0.81         | Adjuvant therapy                   |
|                          | BR                | 21              | 52%             | 183  | 45%             | 2.8                          | 2.8  | 2.8                           | 0.971        | 0.77         |                                    |
|                          | Brain relapse     | 21              | 95%             | 183  | 92%             | NA                           | NA   | NA                            | 0.941        | 0.82         |                                    |
|                          | Lung relapse      | 158             | 75%             | 46   | 87%             | NA                           | 5    | NA                            | 1.045        | 0.78         |                                    |
| GSE19615                 | RFS               | 60              | 90%             | 55   | 84%             | NA                           | NA   | NA                            | 1.021        | 0.94         | Adjuvant anthracycline             |
| GSE20685                 | DSS               | 271             | 73%             | 56   | 84%             | 14                           | NA   | 14.1                          | 1.023        | 0.84         | Adjuvant chemotherapy              |
|                          | MFS               | 292             | 73%             | 35   | 86%             | NA                           | NA   | NA                            | 1.013        | 0.91         |                                    |
| GSE21653                 | DFS               | 126             | 67%             | 140  | 61%             | 3.8                          | NA   | 4                             | 0.886        | 0.33         | Adjuvant chemotherapy              |
| GSE58812                 | MFS               | 67              | 79%             | 40   | 58%             | NA                           | NA   | NA                            | 0.642        | <b>0.043</b> | Adjuvant chemotherapy              |
|                          | OS                | 95              | 76%             | 12   | 50%             | NA                           | NA   | NA                            | 0.76         | 0.21         |                                    |
| GSE25066                 | DRFS              | 149             | 70%             | 359  | 81%             | NA                           | NA   | NA                            | <b>1.297</b> | <b>0.011</b> | Neoadjuvant paclitaxel therapy     |
| Metabric Canadian Cohort | OS                | 120             | 72%             | 874  | 52%             | 12.2                         | 12.4 | 12.4                          | 0.965        | 0.5          | Hormonal, radio- and chemotherapy  |
| Metabric British Cohort  | OS                | 402             | 61%             | 595  | 51%             | 13.2                         | 10.8 | 12.2                          | 0.895        | <b>0.03</b>  | Hormonal, radio- and chemotherapy  |

This trend was most clear in GSE25066 and in this cohort, patients were primarily treated with paclitaxel in a neo-adjuvant setting. Patients with CS/M phenotype had shorter relapse-free survival times when compared with NS/E phenotype (Figure 5.17 A). The cut-off values used for this KM analysis was generated using the log rank multiple cutoff (LRMC) test (Figure 5.17 B).

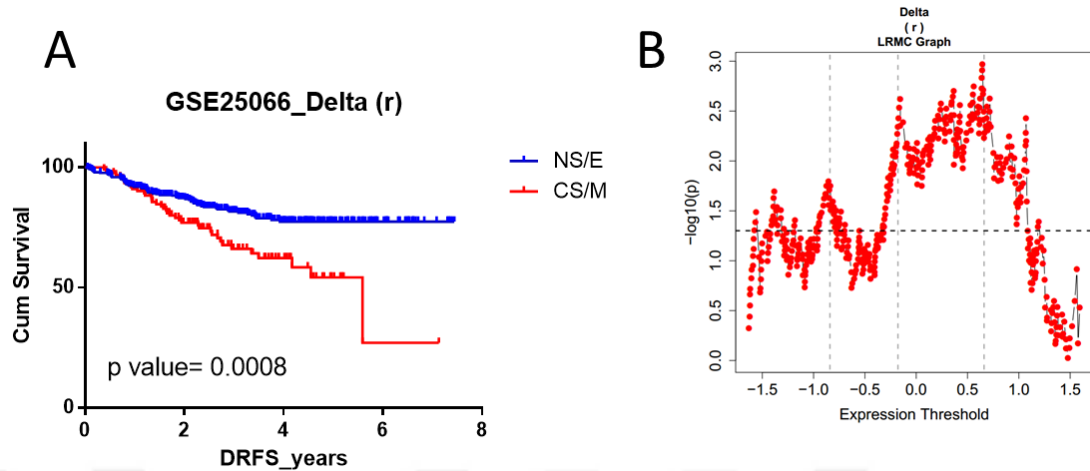


**Figure 5.17: Survival analysis of GSE25066.** (A) CS/M phenotype showed worse distant relapse free survival when compared with NS/E phenotype. (B) 0.6459 was used as a cutoff for dividing patients into CS/M and NS/E subgroups. This was the best cutoff as shown by LRM approach. Figure retrieved from (Akbar et al., 2020a).

In this dataset, delta(r) distributions were significantly different among different molecular subtypes in GSE25066 (Figure 5.18). Nevertheless, CNCL based classification was a significant independent prognostic marker in a multivariate analysis including T stage, N stage and PAM50 subtypes in the same cohort (Table 5.15).

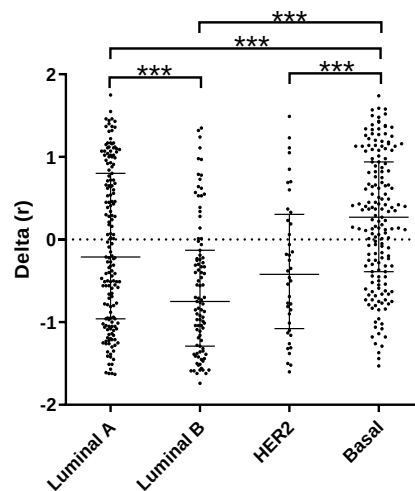


**Figure 5.18: Delta (r) based distribution of breast cancer patients in molecular subtypes in GSE25066.** Figure retrieved from (Akbar et al., 2020a).

**Table 5.15: Multivariate analysis of SS delta (r) with clinical parameters such as T stage, N stage and breast cancer molecular subtypes, showed SS delta (r) as a significant independent prognostic marker.** Table retrieved from (Akbar et al., 2020a)

| Parameters   | Sig.    | HR    |
|--------------|---------|-------|
| SS delta (r) | 0.010   | 1.371 |
| N stage      | 0.0005  | 1.425 |
| T stage      | 0.002   | 1.398 |
| Luminal B    | 0.00003 |       |
| Basal        | 0.027   | 1.906 |
| HER2         | 0.169   | 1.647 |
| Luminal A    | 0.052   | 0.503 |

This analysis clearly shows a strong relation between CNCL identified groups and paclitaxel treatment suggesting CNCL might be beneficial in determining outcomes in patients treated with paclitaxel in a neo-adjuvant setting. Additionally, our previous *in silico* analysis also points in the same direction as cell lines treated with paclitaxel for a longer period, switch from NS/E phenotype to CS/M phenotype.

Lastly, in 22 survival tests performed for multiple cohorts with different endpoints, we did not find any compatible significant association between CNCL and survival end-measures. In light of our very consistent results both *in silico* and *in vitro*, we believe that this lack of association is due to breast cancer heterogeneity and lack of treatment data.

## 5.2. Chemosensitivity markers of paclitaxel, cisplatin and doxorubicin in breast cancer

As we have previously shown *in silico* that CSC genes can predict sensitivity to paclitaxel and doxorubicin among cells treated with these drugs under a prolonged period. So next we asked if the same genes can predict differential sensitivity for triple-negative breast cancer cell lines. We used 4 CSC cancer cell lines (CAL51, MDAMB157, MDAMB231 and MDAMB436) and 7 non-CSC cell lines (BT20, HCC1143, HCC1937, HCC38, HCC70, MDAMB453 and MDAMB468) to identify paclitaxel, cisplatin and doxorubicin sensitivity. Surprisingly, we did not find any significant difference between CSC cytotoxicity and non-CSC cytotoxicity for all 3 drugs (Figure 5.19). Pearson correlation between Delta (r) and *in vitro* cytotoxicity data showed no significant correlation as well

(Paclitaxel (r): -0.02, cisplatin (r): 0.07 and doxorubicin (r): -0.2). This discordance can be explained in terms of innate and acquired resistance. As in our *in silico* analysis, cell lines were treated with respective drugs over a prolonged period, granting these cell stem like properties and drug resistance. Similarly, in the cohort GSE25066 where CSC gene list showed prognostic importance, patients were treated with paclitaxel in neoadjuvant settings which generated CSC like tumors.

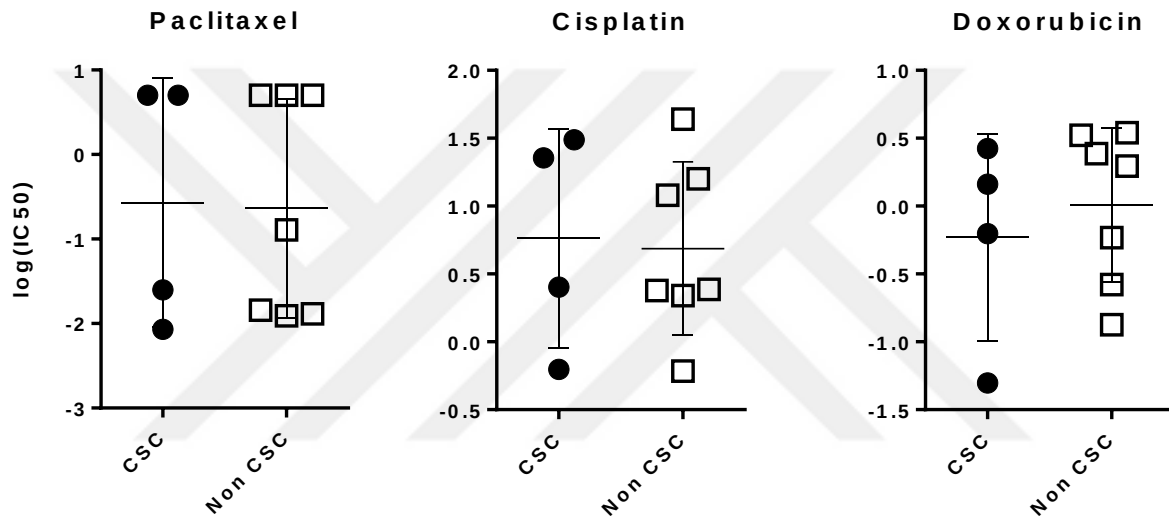


**Figure 5.19: Paclitaxel, cisplatin and doxorubicin cytotoxicity data *in vitro*.** Cytotoxicity comparison of CSC and non-CSC like cancer cell lines for paclitaxel, cisplatin and doxorubicin showed no significant difference *in vitro*.

We found the similar results for docetaxel (paclitaxel), cisplatin and doxorubicin sensitivity for CSC and non-CSC classes in CCLE and CGP datasets (Figure 5.20).

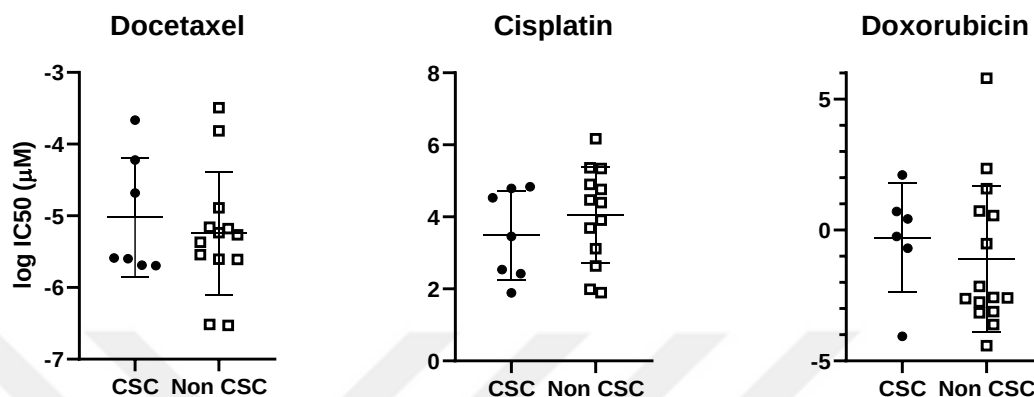


**Figure 5.20: Paclitaxel, cisplatin and doxorubicin cytotoxicity data *in silico*.** Cytotoxicity comparison of CSC and non-CSC like cancer cell lines for docetaxel (paclitaxel), cisplatin and doxorubicin showed no significant difference *in silico*.

CSC gene signature could not predict biomarkers for 3 drugs (paclitaxel, cisplatin and doxorubicin) so we next decided to analyze CGP drug data and our *in vitro* cytotoxicity data for these drugs along with gene expression data (CCLE and CGP). We correlated drug cytotoxicity values with gene expression data and found correlated genes. Then regression equations were derived using these genes to predict cell cytotoxicity *in silico*.

### 5.2.1. Paclitaxel biomarkers

We correlated our *in silico* cytotoxicity data (Activity area) with gene expression data of CCLE and CGP separately. Among correlated genes in either direction, we selected commonly correlated genes (10 genes) between both correlation analyses. Additionally, as gene number was small so we also selected the top 10 up and down correlated genes in both analyses as well. We then used these genes to develop linear regression models using “Best Subset” option of Minitab with *in vitro* drug data as outcome. Models were generated using combinations of 9 genes at a time and analysis was repeated until all combinations were tested. Among resulting models, only those which contained 2 genes and could predict IC<sub>50</sub> (17 models) were considered further for validation in 3 datasets *in silico*. All

models are reported in Table 5.16. Among these, we selected 3 models as they could be validated in at least 2 datasets which are highlighted in green. “Val” and “Un-Val” shows if the model was validated or not in that dataset.

**Table 5.16: Linear regression cytotoxicity models of paclitaxel.**

| Model number | Paclitaxel Activity Area             | Adj. R square | p value | Validation Datasets |          |          |
|--------------|--------------------------------------|---------------|---------|---------------------|----------|----------|
|              |                                      |               |         | GSE12791            | GSE23399 | GSE28784 |
| Model 1      | =411+39.8(AMD1)-87.8(SLC25A28)       | 78.10%        | 0.001   | Val                 | Val      | Un-Val   |
| Model 2      | =139-36.8(TSPO)+44.3(WDR61)          | 73.90%        | 0.002   | Un-Val              | Val      | Val      |
| Model 3      | =-323+15.9(LHX2)+39.8(WDR61)         | 70.20%        | 0.003   | Un-Val              | Val      | Val      |
| Model 4      | =-1061+70.3(AMD1)+71.4(NFYC-AS1)     | 88.70%        | 0       | Un-Val              | Un-Val   | Un-Val   |
| Model 5      | =-896+16.0(LHX2)+74.2(PSMA4)         | 79.70%        | 0.001   | Un-Val              | Val      | Un-Val   |
| Model 6      | =1550-291(LOC101928833)-104(CYP3A43) | 79.70%        | 0.001   | Un-Val              | Un-Val   | Un-Val   |
| Model 7      | =-896+74.2(PSMA4)+16.0(LHX2)         | 79.70%        | 0.001   | Un-Val              | Un-Val   | Un-Val   |
| Model 8      | =645-76.1(SLC25A28)+12.3(LHX2)       | 79.60%        | 0.001   | Un-Val              | Val      | Un-Val   |
| Model 9      | =905-75.4(SLC25A28)-48.7(BANK1)      | 78.90%        | 0.001   | Un-Val              | Val      | Un-Val   |
| Model 10     | =905-75.4(SLC25A28)-48.7(BANK1)      | 78.90%        | 0.001   | Un-Val              | Val      | Un-Val   |
| Model 11     | =-532+72.8(PSMA4)-63.4(BANK1)        | 78.70%        | 0.001   | Un-Val              | Un-Val   | Un-Val   |
| Model 12     | =-532-63.4(BANK1)+72.8(PSMA4)        | 78.70%        | 0.001   | Un-Val              | Un-Val   | Un-Val   |
| Model 14     | =-356+46.5(IREB2)+12.5(LHX2)         | 72.60%        | 0.002   | Un-Val              | Un-Val   | Un-Val   |
| Model 15     | =733-59.9(BANK1)-88.6(KCNIP2-AS1)    | 71.90%        | 0.003   | Un-Val              | Un-Val   | Un-Val   |
| Model 16     | =725-57.0(BANK1)-88.0(C10orf12)      | 70.70%        | 0.003   | Un-Val              | Un-Val   | Val      |
| Model 17     | =-276-69.8(BANK1)+82.6(PIGB)         | 63.80%        | 0.007   | Un-Val              | Un-Val   | Un-Val   |

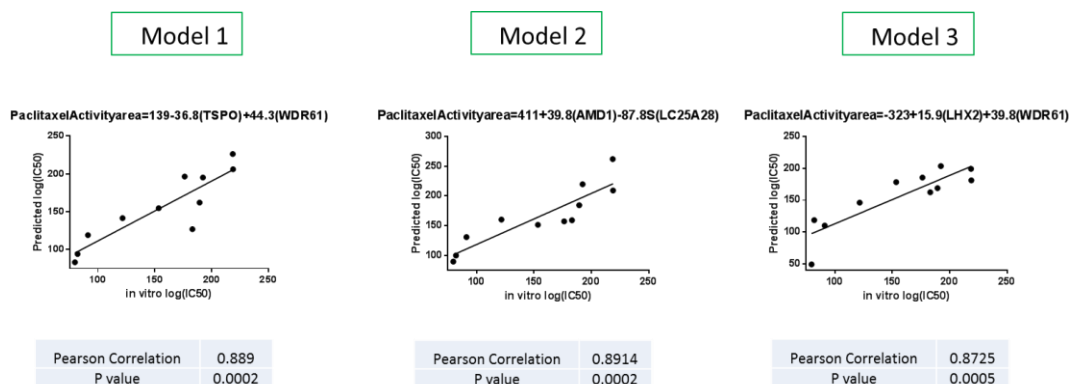


**Figure 5.21: Pearson correlation analysis between paclitaxel predicted IC50 and *in vitro* cytotoxicity data for selected 3 models.** Predicted IC50 of paclitaxel showed a high correlation with *in vitro* cytotoxicity data for Model 1, 2 and 3.

For validation of linear regression models, we used GSE12791 (Luo et al., 2011), GSE23399 (Chan et al., 2016) and GSE28784 as validation datasets. GSE12791 contains the gene expression data for 4 groups: 1: MDAMB231 cells made resistant to paclitaxel upon 3 days treatment followed by 7 days exposure to control media (MDA-PR). 2: MDA-PR cells further treated with 30nM paclitaxel for 3 days followed by 7 days of exposure to control media (MDA-PR-Paclitaxel). 3: MDA-PR cells treated with 1  $\mu$ M Bexarotene for 10 days (MDA-PR-Paclitaxel-Bexarotene). 4: Parent cells not treated with any drug. Only Model 1 could be validated in this dataset as shown in Figure 5.22 as predicted activity area was lower for MDA-PR and MDA-PR-Paclitaxel compared with MDA-Parental and MDA-PR-Paclitaxel-Bexarotene.

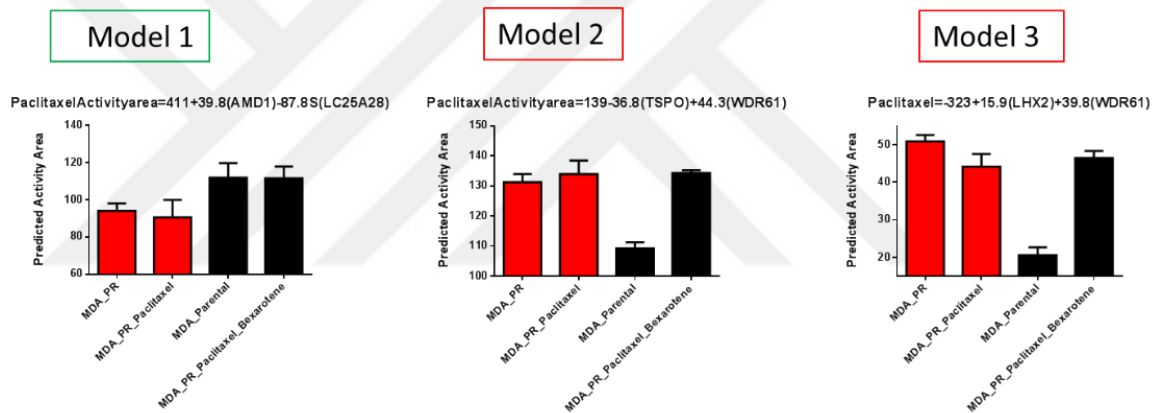


**Figure 5.22: Validation of 3 paclitaxel models in dataset GSE12791.** Model 1 differentiated paclitaxel resistant groups (MDA-PR and MDA-PR-Paclitaxel) from paclitaxel sensitive groups (MDA-Parental and MDA-PR-Paclitaxel-Bexarotene). Model2 and 3 could not be validated.

In dataset GSE23399, Breast cancer associated fibroblasts were treated with paclitaxel and gene expression was measured at day1, 3 and 7. All 3 models could successfully predict higher activity area for cells collected at 1 and 3 days along with control cells when compared with cells collected at day 7 (Figure 5.23).

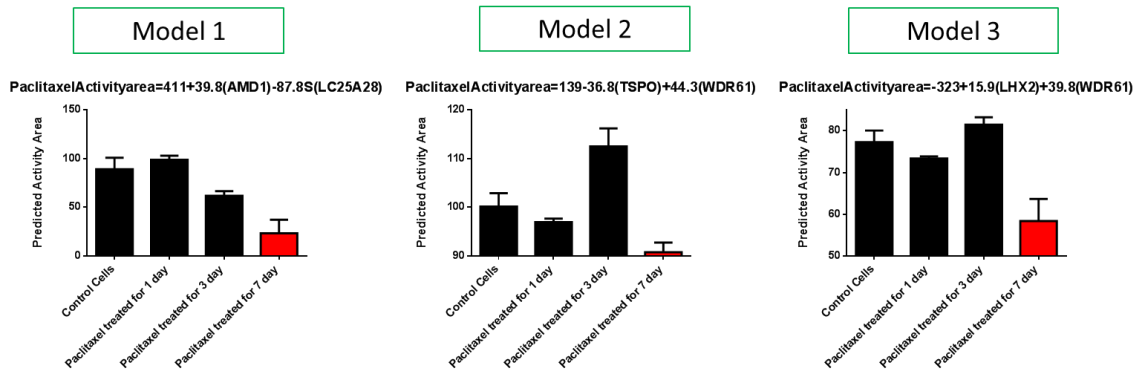


**Figure 5.23: Validation of 3 paclitaxel models in dataset GSE23399.** All 3 models predicted higher activity area for control cells, paclitaxel treated cells collected at 1 and 3 days (sensitive cells) when compared with cells collected at 7 days (resistant cells).

In dataset GSE28784, MDAMB231 cells were either treated with paclitaxel (resistant cells) or verapamil which is an inhibitor of drug efflux pump proteins (sensitive cells). When we applied our models to these cells, both Model 2 and 3 predicted higher activity area for cells treated with verapamil (sensitive cells) when compared with the paclitaxel treated group (resistant group) as shown in Figure 5.24.

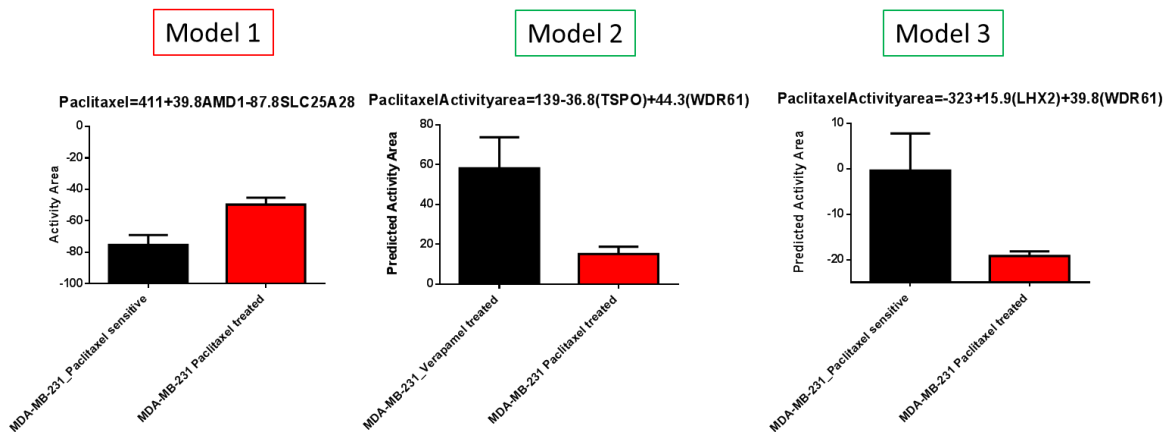


**Figure 5.24: Validation of 3 paclitaxel models in dataset GSE28784.** Model 2 and 3 predicted higher activity area for cells treated with verapamil (sensitive cells) when compared with cells treated with paclitaxel (resistant cells).

In conclusion, our 2 gene linear regression models have the potential to predict paclitaxel drug response in breast cancer cell lines. These models will be further validated both *in silico* and *in vitro* (using cell lines which were not part of our discovery dataset).

### 5.2.2. Cisplatin biomarkers

We correlated our *in silico* cytotoxicity data (IC50) and CGP cytotoxicity data with gene expression data of CGP separately. Among correlated genes in either direction, we selected commonly correlated genes (16 genes) between both correlation analyses. We then used these genes to develop linear regression models using “Best Subset” option of Minitab with *in vitro* drug data as outcome. Models were generated using combinations of 9 genes at a time and analysis was repeated until all combinations were tested. Among resulting models, only those which contained 2 genes and could predict IC50 (9 models) were considered further for validation *in silico*. All models are reported in Table 5.17. Among these, we selected 2 models as they could be validated in at least 2 datasets (in vitro data and GSE28274 (Alborzinia et al., 2011)) which are highlighted in green. “Val” and “Un-Val” shows if the model was validated or not in that dataset.

**Table 5.17: Linear regression cytotoxicity models of cisplatin.**

| Model No | Cisplatin Models (Log10(IC50))        | Adj R square | p value | <i>in vitro</i> data | GSE28274 |
|----------|---------------------------------------|--------------|---------|----------------------|----------|
| Model 1  | =-3.10+0.343(C14orf169)+0.388(DOHH)   | 47.30%       | 0.003   | Val                  | Val      |
| Model 2  | =0.59+0.326(C14orf169)-0.159(RBBP4)   | 42.30%       | 0.006   | Val                  | Val      |
| Model 3  | =-3.68+0.284(C14orf169)+0.351(DDX27)  | 46.10%       | 0.004   | Val                  | Un-Val   |
| Model 4  | =0.76+0.403(C14orf169)-0.226(DHX9)    | 46.90%       | 0.003   | Val                  | Un-Val   |
| Model 5  | =-3.71+0.241(SREK1)+0.315(DDX27)      | 45.50%       | 0.004   | Val                  | Un-Val   |
| Model 6  | =-3.19+0.335(DDX27)+0.207(GTF2F1)     | 45.10%       | 0.004   | Val                  | Un-Val   |
| Model 7  | =0.201+0.333(C14orf169)-0.132(PGRMC1) | 45.00%       | 0.004   | Val                  | Un-Val   |
| Model 8  | =0.97-0.140(RBBP4)+0.219(GTF2F1)      | 37.20%       | 0.012   | Val                  | Un-Val   |
| Model 9  | =1.14-0.177(DHX9)+0.249(GTF2F1)       | 36.70%       | 0.013   | Val                  | Un-Val   |

As expected, our *in vitro* cytotoxicity data showed a high correlation with predicted IC50 values as shown in Figure 5.25 for selected 2 models.

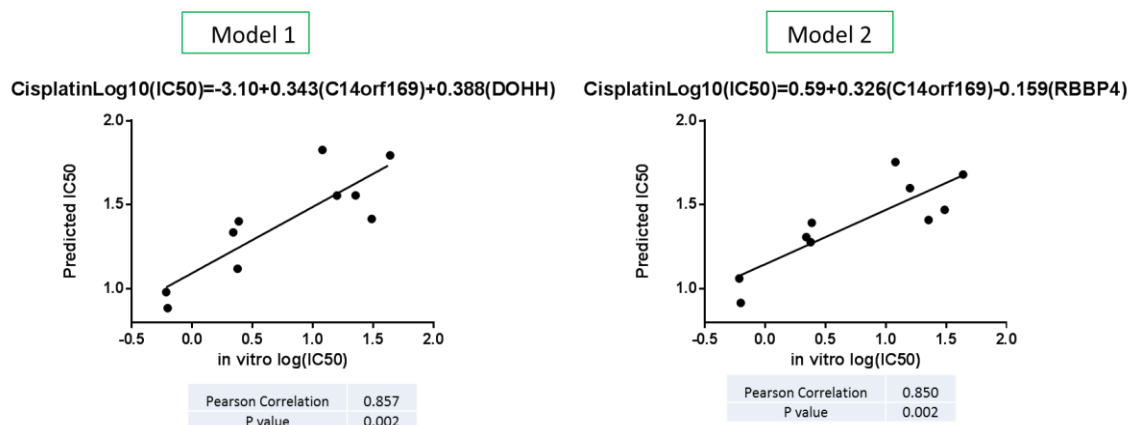


**Figure 5.25: Validation of 2 cisplatin models with *in vitro* cisplatin cytotoxicity data.** Predicted IC<sub>50</sub> of cisplatin showed a high correlation with *in vitro* cytotoxicity data for both Model 1 and 2.

Selected models were also validated in GSE28274 (Alborzinia et al., 2011). In this dataset, MCF7 cells were treated with 50  $\mu\text{M}$  cisplatin. Cells were collected after 8- and 10-hours treatment. Both selected models showed higher IC<sub>50</sub> predicted value for treated cells when compared with control cells as shown in Figure 5.26.

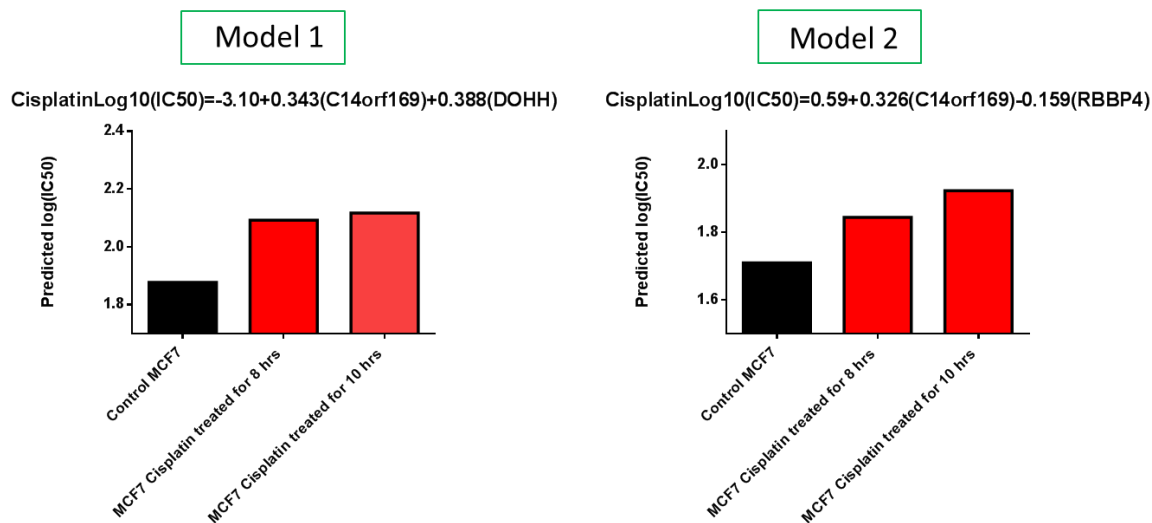


**Figure 5.26: Validation of 2 cisplatin models in dataset GSE28274.** Both cisplatin models predicted Lower IC<sub>50</sub> for MCF7 control cells (sensitive cells) when compared with cells collected after 8- and 10-hours cisplatin treatment (resistant cells).

In conclusion, our 2 gene linear regression models have the potential to predict cisplatin drug response in breast cancer cell lines. These models will be further validated both *in silico* and *in vitro* (using cell lines which were not part of our discovery dataset).

### 5.2.3. Doxorubicin biomarkers

We correlated our *in silico* cytotoxicity data (IC50) and CGP cytotoxicity data with gene expression data of CGP and CCLE individually. Among correlated genes in either direction, we selected commonly correlated genes (19 genes) between all 4 correlation analyses. We then used these genes to develop linear regression models using “Best Subset” option of Minitab with *in vitro* drug data as outcome. Models were generated using combinations of 9 genes at a time and analysis was repeated until all combinations were tested. Among resulting models, only those which contained 2 genes and could predict IC50 (10 models) were considered further for validation *in silico*. All models are reported in Table 5.18. Among these, we selected 2 models as they could be validated in at least 4 datasets which are highlighted in green. “Val” and “Un-Val” shows if the model was validated or not in that dataset.

**Table 5.18: Linear regression cytotoxicity models of doxorubicin**

| Model No | Doxorubicin                                     | Adj. R square | P value | in vitro data | GSE3929 | GSE13477 | GSE24460 | GSE72362 |
|----------|-------------------------------------------------|---------------|---------|---------------|---------|----------|----------|----------|
| Model 1  | =10.1-0.826(FAM136A//LOC100287852)-0.663(NME5)  | 50.80%        | 0.001   | Val           | Val     | Val      | Val      | Val      |
| Model 2  | =3.92-0.220(SCGB1D2)-0.694(FZD3)                | 40.00%        | 0.005   | Val           | Val     | Un-Val   | Val      | Val      |
| Model 3  | =-3.90+0.729(STX4)-0.548(NME5)                  | 51.80%        | 0.001   | Val           | Un-Val  | Val      | Un-Val   | Val      |
| Model 4  | =-3.90-0.548(NME5)+0.729(STX4)                  | 51.80%        | 0.001   | Val           | Un-Val  | Un-Val   | Un-Val   | Val      |
| Model 5  | =-1.76+0.547(STX4)-0.760(PTPRD)                 | 46.80%        | 0.002   | Val           | Un-Val  | Un-Val   | Un-Val   | Un-Val   |
| Model 6  | =8.35-0.556(FAM136A//LOC100287852)-0.905(PTPRD) | 45.40%        | 0.002   | Val           | Un-Val  | Un-Val   | Un-Val   | Un-Val   |
| Model 7  | =8.35-0.556(FAM136A//LOC100287852)-0.905(PTPRD) | 45.40%        | 0.002   | Val           | Un-Val  | Un-Val   | Un-Val   | Un-Val   |
| Model 8  | =-11.3+0.738(STX4)+0.439(RAB2A)                 | 42.10%        | 0.004   | Val           | Un-Val  | Un-Val   | Val      | Un-Val   |
| Model 9  | =-4.84-0.649(FZD3)+0.662(EEF1D)                 | 42.10%        | 0.004   | Val           | Val     | Un-Val   | Val      | Un-Val   |
| Model 10 | =-11.2+0.696(STX4)+0.428(EEF1D)                 | 41.40%        | 0.004   | Val           | Un-Val  | Un-Val   | Val      | Un-Val   |

As expected, our *in vitro* cytotoxicity data showed a high correlation with predicted IC50 values as shown in Figure 5.27 for selected 2 models.

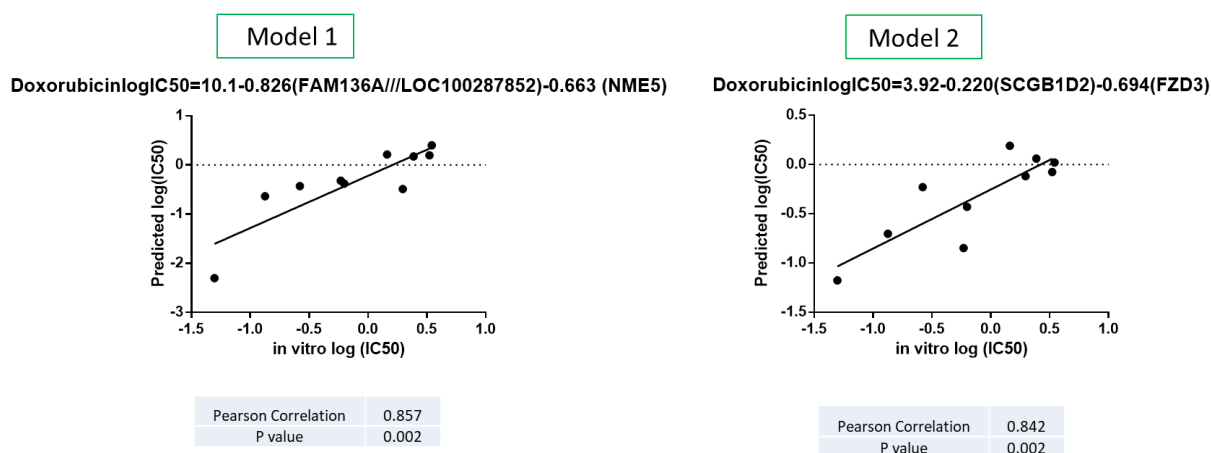


**Figure 5.27: Validation of 2 doxorubicin models with *in vitro* cytotoxicity data.** Predicted IC50 of doxorubicin showed a high correlation with *in vitro* cytotoxicity data for both Model 1 and 2.

Doxorubicin models were then validated in dataset GSE13477 (Stockwin et al., 2009). In this dataset, MCF7 cells were treated with doxorubicin for 24 hours. Both models predicted higher IC50 for treated cells when compared with control cells as expected (Figure 5.28).

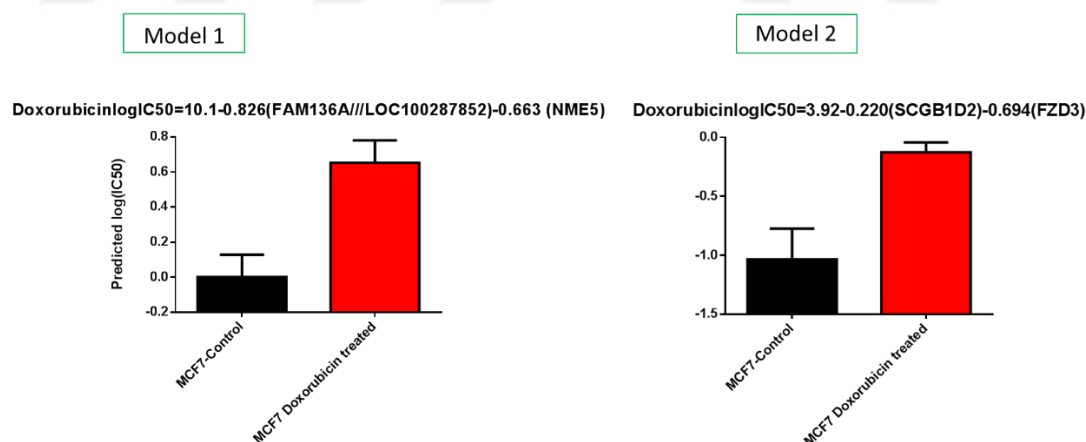


**Figure 5.28: Validation of 2 doxorubicin models in dataset GSE13477.** Both doxorubicin models predicted Lower IC50 for MCF7 control cells (sensitive cells) when compared with cells treated with doxorubicin for 24 hours (resistant cells).

Selected models were next validated in GSE24460 (Calcagno et al., 2010). This dataset contains expression data for MCF7 cells which were cultured in high dose doxorubicin (860 nM) every other passage (MCF-7 ADR multistep doxorubicin selected) and MCF7

control cells. Both models predicted higher IC50 for treated cells when compared with control cells as expected (Figure 5.29).

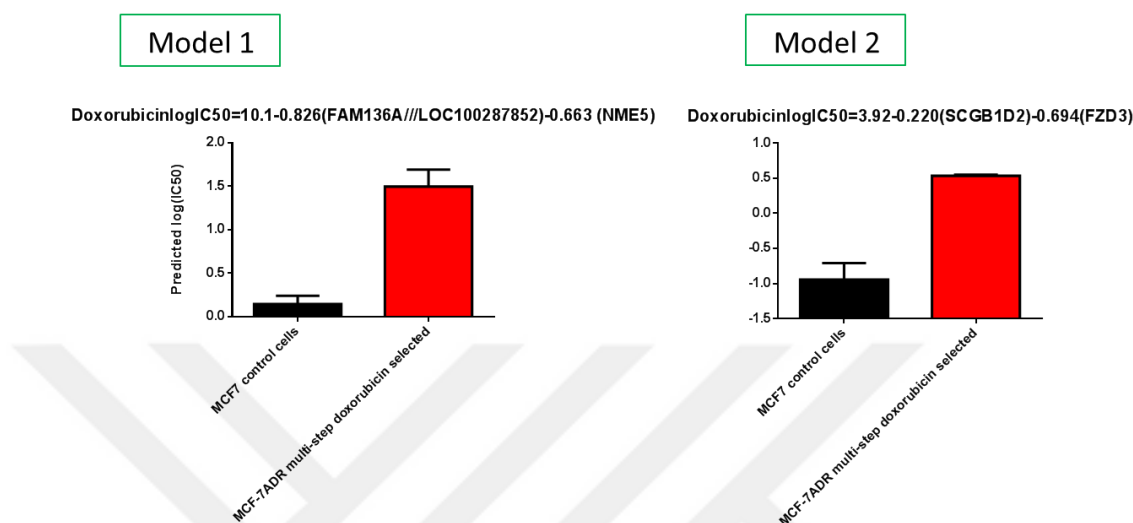


**Figure 5.29: Validation of 2 doxorubicin models in dataset GSE24460.** Both doxorubicin models predicted Lower IC50 for MCF7 control cells (sensitive cells) when compared with cells treated with doxorubicin (MCF-7 ADR multistep doxorubicin selected).

Selected models were next validated in GSE72362 (O'Brien et al., 2015). In this dataset, SKBR3 cells were treated with 90 nM epirubicin (epimer of doxorubicin) for 3 days. These treated cells were FACS sorted in CD44 high and low. The expression of these cells was then compared with parental SKBR3 untreated cells. Among epirubicin treated samples, both models predicted higher IC50 for CD44 high cells when compared with CD44 low cells. And untreated cells showed the least IC50 as expected (Figure 5.30).

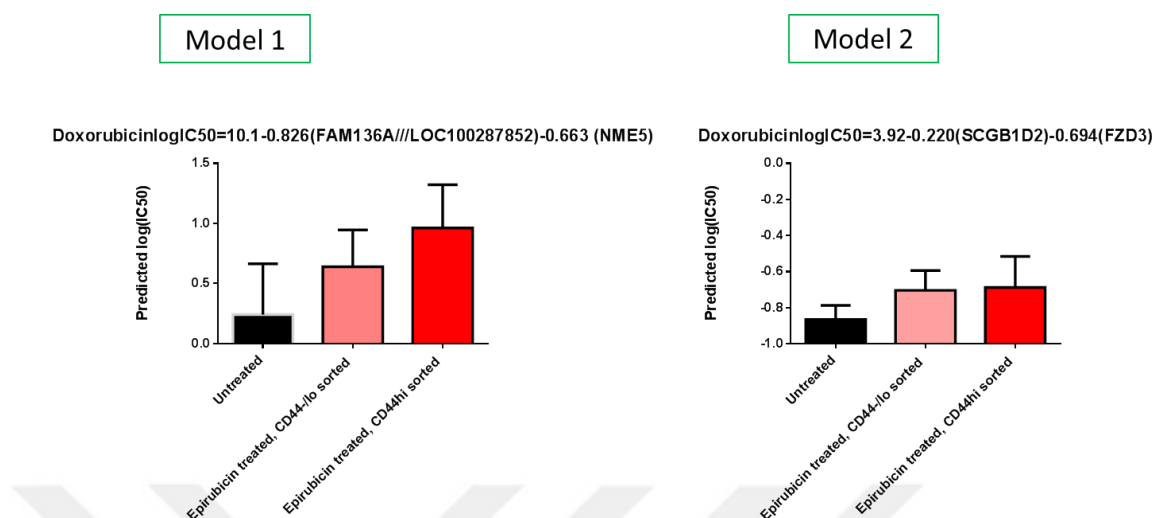


**Figure 5.30: Validation of 2 doxorubicin models in dataset GSE72362.** Both doxorubicin models predicted Lower IC<sub>50</sub> for SKBR3 control cells (sensitive cells) when compared with cells treated with epirubicin. Among epirubicin treated samples, predicted IC<sub>50</sub> was higher for CD44 high cells when compared with CD44 low cells.

Selected models were next validated in GSE3929 (Munkácsy et al., 2010). In this dataset, resistant MCF7 models were developed using long term treatment. Furthermore, additional cells were treated with 0.05 µg/ml of doxorubicin for 24 hours as well. Our model predicted higher IC<sub>50</sub> for multistep resistant models when compared with cells treated for 24 hours or untreated parental cells (Figure 5.31).

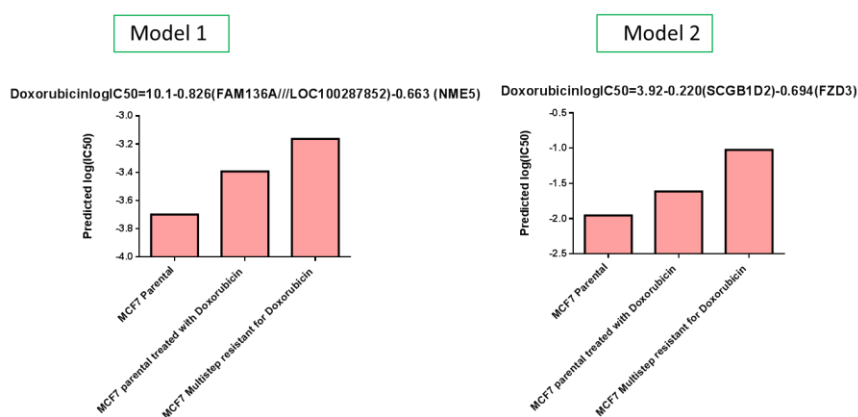


**Figure 5.31: Validation of 2 doxorubicin models in dataset GSE3929.** MCF7 cells were treated with doxorubicin for 24 hours and for a longer period to develop resistant cells. Both short term treated and untreated parental cells predicted IC<sub>50</sub>s were lower when compared to multistep resistant cells for both models.

In conclusion, our 2 gene linear regression models have the potential to predict doxorubicin drug response in breast cancer cell lines. These models will be further validated both *in silico* and *in vitro* (using cell lines which were not part of our discovery dataset).

### **5.3. Breast cancer plasticity**

Plasticity is the phenomenon where breast cancer cells switch from one class to another in response to either metastasis, microenvironment, or therapy. Three types of plasticity exist in breast cancer: PAM50 plasticity, IHC plasticity and EMT plasticity.

#### **5.3.1. PAM50 plasticity**

To study PAM50 plasticity, we analyzed different datasets that contained matched gene expression data for breast cancer samples, before or after therapy and primary or metastatic tissue.

PAM50 classification is dependent on the biopsy collection method as various samples show different classification when different biopsy techniques are used to collect breast cancer samples. In dataset GSE32518 (Gonzalez-Angulo et al., 2012), patients matching samples were collected through fine needle aspiration and core needle biopsy. In total 37 samples were analyzed. When these samples were subtype classified by using PAM50, 78% (29/37) samples were classified in the same classes when both biopsy methods were compared (Table 5.19). Discordant subtype classification was seen for 22% (8/37) samples. Among samples that were identified as HER2 in core needle biopsy, one was identified as Basal and another one as LumB. Only one sample which was identified as LumA by core needle biopsy was identified as LumB by fine-needle aspiration. Normal subtype (identified by core needle biopsy) showed the most discordance as 4 samples were subtyped as Basal and one sample was classified as LumA (fine needle aspiration)

**Table 5.19: Comparison of PAM50 subtyping for samples collected through fine needle aspiration and core needle biopsy (GSE32518). Table retrieved from (Akbar et al., 2020b).**

|                      |        | Fine needle aspiration |      |      |      |        |
|----------------------|--------|------------------------|------|------|------|--------|
|                      |        | Basal                  | HER2 | LumA | LumB | Normal |
| Core needle biopsies | Basal  | 7                      | 0    | 0    | 0    | 0      |
|                      | HER2   | 1                      | 5    | 0    | 1    | 0      |
|                      | LumA   | 0                      | 0    | 9    | 1    | 0      |
|                      | LumB   | 0                      | 0    | 0    | 7    | 0      |
|                      | Normal | 4                      | 0    | 1    | 0    | 1      |

PAM50 plasticity is observed in breast cancer patients which go through chemotherapy as some samples switch from one subtype to another. We analyzed GSE10281 (Creighton et al., 2009) which contained expression data for patient's matched samples before and after Letrozole therapy. In total 18 samples were compared. PAM50 subtype remained the same for 39% (7/18) samples and 61% (11/18) switched their subtype (Table 5.20). Basal (n=2) and LumB (n=1) subtype remained consistent through drug treatment. Among the samples which were HER2 subtype pre-therapy, 1 changed to Basal, 2 changed to LumB and 1 remained unchanged. For LumA pre-therapy samples, 3 switched to LumB and 2 remained unchanged. For samples that were subtyped as Normal, 3 switched to Basal, 1 to LumA, 1 to LumB and 1 did not show any switch. Additionally, the authors reported that residual tumors after chemotherapy switched to a mesenchymal phenotype as expected.

**Table 5.20: Comparison of PAM50 subtypes pre- and post-Letrozole therapy (GSE10281). Table retrieved from (Akbar et al., 2020b).**

|             |        | Post-therapy |      |      |      |        |
|-------------|--------|--------------|------|------|------|--------|
|             |        | Basal        | HER2 | LumA | LumB | Normal |
| Pre-therapy | Basal  | 2            | 0    | 0    | 0    | 0      |
|             | HER2   | 1            | 1    | 0    | 2    | 0      |
|             | LumA   | 0            | 0    | 2    | 3    | 0      |
|             | LumB   | 0            | 0    | 0    | 1    | 0      |
|             | Normal | 3            | 0    | 1    | 1    | 1      |

Nex, we analyzed dataset GSE28844 (Vera-Ramirez et al., 2013) which contained matched pre- and post-therapy expression data for breast cancer patients treated with neoadjuvant anthracyclines and taxanes. PAM50 subtype remained the same for 36% (10/28) samples (Table 5.21). 2 samples basal pre-therapy subtype switched to Normal subtype post-therapy. Similarly, among HER2 pre-therapy subtype, 3 samples switched their subtype to Basal, LumB and Normal. Among pre-therapy LumA subtype samples, 4 samples switched to Normal subtype post-therapy. In this dataset, LumB pre-therapy showed the highest plasticity as 1 sample switched to Basal, 4 switched to LumA and 3 switched to Normal subtype post-therapy. Lastly, only one normal sample switched to LumA. Importantly in this dataset, most samples switched to Normal subtype post-therapy.

**Table 5.21: Comparison of PAM50 subtypes pre- and post-Anthracyclines and Taxanes based neoadjuvant chemotherapy (GSE28844). Table retrieved from (Akbar et al., 2020b).**

|             |        | Post-therapy |      |      |      |        |
|-------------|--------|--------------|------|------|------|--------|
|             |        | Basal        | HER2 | LumA | LumB | Normal |
| Pre-therapy | Basal  | 2            | 0    | 0    | 0    | 2      |
|             | HER2   | 1            | 1    | 0    | 1    | 1      |
|             | LumA   | 0            | 0    | 3    | 0    | 4      |
|             | LumB   | 1            | 0    | 4    | 3    | 3      |
|             | Normal | 0            | 0    | 1    | 0    | 1      |

Next, we analyzed dataset GSE32072 (Gonzalez-Angulo et al., 2012) which contained samples treated with neoadjuvant systemic therapy. In this dataset, 9 patients were treated with anthracycline based therapy, 8 were treated with anthracycline/taxane based therapy and 2 were treated with trastuzumab therapy. This dataset additionally contained 4 samples that were not treated with neoadjuvant chemotherapy and all of these samples retained their phenotype as expected. Among treated samples, 76% (16/21) samples kept their phenotype when matched pre- and post-therapy samples were compared. Only one sample switched from LumA (pre-Therapy) to LumB (Post-therapy) and among normal samples (pre-therapy), 2 switched to Basal, 1 switched to LumA and 1 to LumB post-therapy Table 5.22.

**Table 5.22: Comparison of PAM50 subtypes pre- and post- neoadjuvant chemotherapy (GSE32072). Table retrieved from (Akbar et al., 2020b).**

|             |        | Post-therapy |      |      |      |        |
|-------------|--------|--------------|------|------|------|--------|
|             |        | Basal        | HER2 | LumA | LumB | Normal |
| Pre-therapy | Basal  | 5            | 0    | 0    | 0    | 0      |
|             | HER2   | 0            | 4    | 0    | 0    | 0      |
|             | LumA   | 0            | 0    | 5    | 1    | 0      |
|             | LumB   | 0            | 0    | 0    | 1    | 0      |
|             | Normal | 2            | 1    | 1    | 0    | 1      |

Breast cancer cells when metastasized, can switch to another subtype when compared with their primary tumors. Next, we analyzed matched primary and brain metastasized tissue subtypes in dataset GSE125989 (Iwamoto et al., 2019). This dataset contained 16 patient samples. 50% (8/16) showed the same subtype for primary and metastatic tissues (Table 5.23). Among the samples which switched their phenotype, one basal sample (primary tumor) switched to HER2 subtype (metastatic). 1 HER2 sample switched to Basal and another one switched to Normal subtype. All the normal samples switched to another subtype upon metastasis as 3 samples switched to Basal and 2 to HER2 subtype.

**Table 5.23: Comparison of PAM50 subtyping for primary tissues vs. brain metastatic tissues (GSE125989). Table retrieved from (Akbar et al., 2020b).**

|                |        | Metastasized tissue |      |      |      |        |
|----------------|--------|---------------------|------|------|------|--------|
|                |        | Basal               | HER2 | LumA | LumB | Normal |
| Primary tissue | Basal  | 2                   | 1    | 0    | 0    | 0      |
|                | HER2   | 1                   | 2    | 0    | 0    | 1      |
|                | LumA   | 0                   | 0    | 4    | 0    | 0      |
|                | LumB   | 0                   | 0    | 0    | 0    | 0      |
|                | Normal | 3                   | 2    | 0    | 0    | 0      |

Dataset GSE110590 (Siegel et al., 2018) contains expression data for 16 primary tumor samples and their multi-systemic metastasis to various tissues including brain, liver, lung, skin and pancreases among others. In total 67 different metastatic tissues were evaluated for gene expression from 16 patients. 55% (37/67) metastatic samples showed the same phenotype as of their primary tissue (Table 5.24). Basal subtype was mostly stable as only 2 (sample A15 and A1) samples switched to Normal subtype upon metastasis and the rest retained their subtype in metastatic tissue. Importantly out of 27 metastatic events that happened for 6 patients, 23 metastatic tissues showed the same phenotype when compared with primary tissues. For Her2 subtype, all 4 metastatic tissues showed the same phenotype as of their primary tissue. LumA samples showed the highest level of switching as 78% (21/31) metastatic tissues switched their subtype. Similarly, for one Normal sample, its 5 metastatic growths switched to Basal subtype.

**Table 5.24: Comparison of PAM50 subtyping for primary tissues vs. metastatic tissues (GSE110590).** Table retrieved from (Akbar et al., 2020b).

| Sample name | Primary tissue | PAM50 metastasis |      |      |      |        |
|-------------|----------------|------------------|------|------|------|--------|
|             |                | Basal            | HER2 | LumA | LumB | Normal |
| A11         | Basal          | 5                | 0    | 0    | 0    | 0      |
| A15         | Basal          | 2                | 0    | 0    | 0    | 3      |
| A1          | Basal          | 4                | 0    | 0    | 0    | 1      |
| A23         | Basal          | 5                | 0    | 0    | 0    | 0      |
| A5          | Basal          | 2                | 0    | 0    | 0    | 0      |
| A7          | Basal          | 5                | 0    | 0    | 0    | 0      |
| A8          | HER2           | 0                | 4    | 0    | 0    | 0      |
| A12         | LumA           | 0                | 0    | 2    | 4    | 0      |
| A17         | LumA           | 2                | 0    | 0    | 0    | 0      |
| A26         | LumA           | 3                | 0    | 0    | 0    | 0      |
| A28         | LumA           | 0                | 0    | 3    | 3    | 0      |
| A2          | LumA           | 0                | 1    | 0    | 2    | 0      |
| A30         | LumA           | 5                | 0    | 0    | 0    | 0      |
| A34         | LumA           | 0                | 0    | 4    | 0    | 0      |
| A4          | LumA           | 0                | 1    | 1    | 0    | 0      |
| A20         | Normal         | 5                | 0    | 0    | 0    | 0      |

In conclusion, we show that some breast tumors switch their molecular subtype upon drug treatment or metastasis. As breast cancer patients are treated based on their subtype, so we recommend reevaluating the subtype of the non-responsive patient or upon metastasis to alter the treatment plan if necessary, based upon the subtype switch.

#### 5.4. Metastatic biomarkers

To identify metastatic biomarkers, 7 patient datasets with metastasis free survival were used. Cohort treatment protocol and patient numbers are mentioned in Table 5.25.

**Table 5.25: Patient number and treatment protocols for cohorts used for metastasis biomarker analysis.** Table retrieved from (Akbar et al., 2020b).

| Dataset               | Patient number | Treatment protocol            | References              |
|-----------------------|----------------|-------------------------------|-------------------------|
| GSE2603               | 82             | Treatment not disclosed       | (Bos et al., 2009)      |
| GSE6532 (U 133 A+B)   | 293            | Tamoxifen treated             | (Loi et al., 2007)      |
| GSE6532 (U133 plus 2) | 87             | Tamoxifen treated             | (Loi et al., 2007)      |
| GSE7390               | 198            | Systemically untreated        | (Desmedt et al., 2007)  |
| GSE11121              | 200            | Only surgery was performed    | (Schmidt et al., 2008)  |
| GSE12276              | 204            | Adjuvant chemotherapy treated | (Bos et al., 2009)      |
| GSE20685              | 327            | Adjuvant chemotherapy treated | (Kao et al., 2011)      |
| GSE58812              | 108            | Adjuvant chemotherapy treated | (Jezequel et al., 2015) |

We performed Cox regression hazard analysis for all datasets. Only those genes were selected as metastatic biomarkers which showed the same direction (good prognosis with  $HR < 1$  or bad prognosis with  $HR > 1$ ) and were significant for at least 5 analyses, were selected. In total 40 probesets passed the selection criteria. These probesets were then ranked by their cox p value in each dataset and rank sum analysis was performed by adding the ranks of probesets in each dataset separately for good and bad prognostic biomarkers. From these top 10 good and bad prognostic biomarkers were selected with the least rank sum score. Additionally, for genes that have 2 or more probesets and showed least rank sum scores for those probesets, only one probeset with the least rank sum score was selected. All selected genes are shown in Figure 5.32.

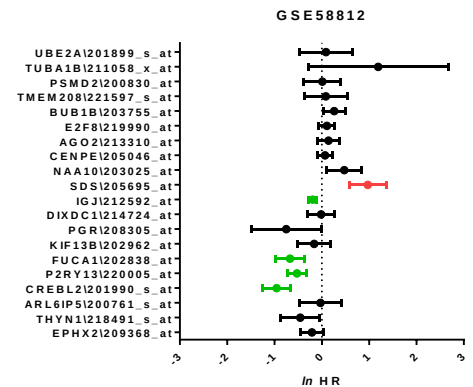
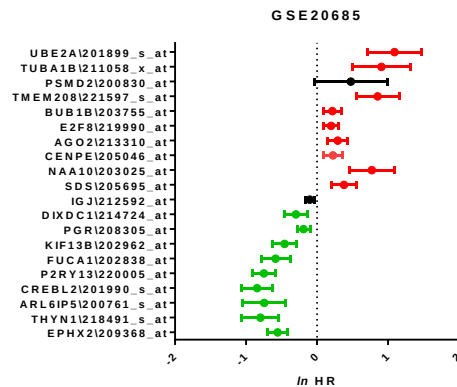
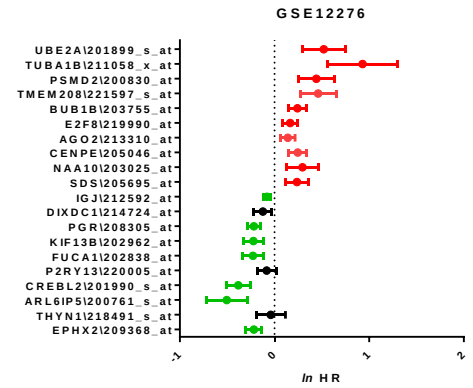
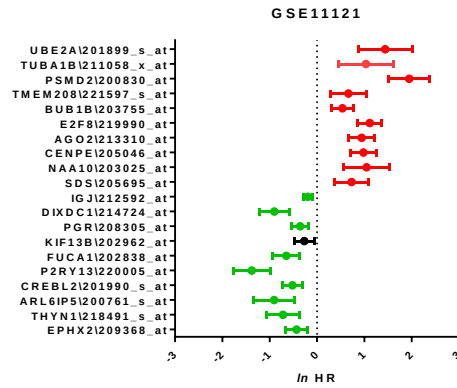
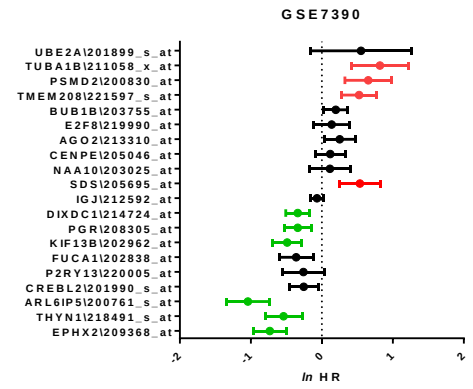
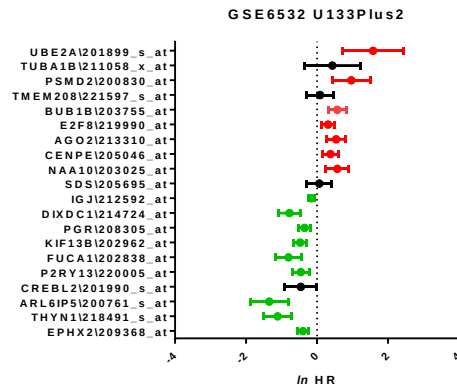
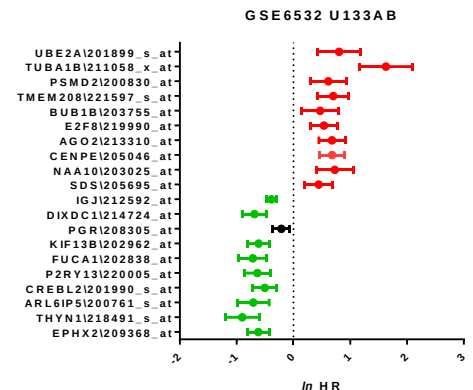
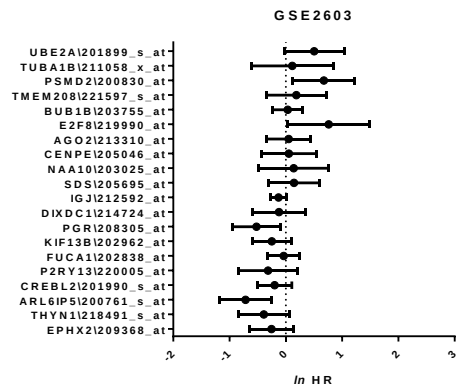


**Figure 5.32: Biomarkers for breast cancer metastasis:** 20 genes were selected as biomarkers for breast cancer metastasis. Ln HR shows natural logarithm for hazard ration on the x-axis. Y-axis shows gene names and probe IDs. If good prognostic gene was significant, it is shown in green while bad prognostic significant genes are shown in red. Black bars show insignificant genes in each dataset. Figure retrieved from (Akbar et al., 2020b).

These identified genes will be further validated *in silico* and *ex vivo* in future studies.

#### **5.4.1. Prognostic importance of hematopoietic cells for metastasis**

Different immune cells play important role in the metastasis cascade. To evaluate this phenomenon, we used Cibersort tool. This tool can quantify relative levels of 22 mature hematopoietic populations and activation states in human cancer using gene expression data. We used this tool in relative quantification of 22 different cells for each tumor in 7 cohorts with metastasis free survival. After cell quantification, these values were used to calculate HRs and cox p value to evaluate the prognostic importance of these cells for metastasis. Table 5.26 and Table 5.27 show HR and cox p values for all cells whose relative quantification was performed.

**Table 5.26: Hazard ratios (HR) of hematopoietic cells for metastasis free survival. Cells highlighted in green show good prognosis and red shows bad prognosis. Cells with thick borders show HR with significant Cox p value.**

|                              | GSE2603  | GSE6532<br>AB | GSE6532<br>plus 2 | GSE7390 | GSE11121 | GSE12276 | GSE20685 | GSE58812 |
|------------------------------|----------|---------------|-------------------|---------|----------|----------|----------|----------|
| Cell name                    | exp(HR)  | exp(HR)       | exp(HR)           | exp(HR) | exp(HR)  | exp(HR)  | exp(HR)  | exp(HR)  |
| B cells naive                | 0.002    | 3.627         | 0.000             | 102.631 | 1.449    | 0.039    | 0.022    | 0.003    |
| B cells memory               | 0.013    | 0.175         | 8.046             | 0.016   | 0.629    | 2.268    | 0.004    | 0.001    |
| Plasma cells                 | 4.313    | 0.301         | 0.338             | 0.522   | 0.053    | 0.342    | 10.801   | 2.312    |
| T cells CD8                  | 0.100    | 2.489         | 0.753             | 0.789   | 1.110    | 2.271    | 0.010    | 0.033    |
| T cells CD4 naive            | 0.000    | 1.859         | 1.888             | 0.000   | 0.005    | 0.017    | 7.079    | 0.000    |
| T cells CD4 memory resting   | 2929.373 | 0.001         | 6.869             | 103.642 | 2.008    | 0.198    | 6.947    | 787.182  |
| T cells CD4 memory activated | 0.000    | 0.000         | 6.54E+07          | 0.003   | 0.000    | 0.003    | 0.000    | 0.000    |
| T cells follicular helper    | 0.301    | 0.003         | 0.235             | 0.002   | 0.003    | 56.642   | 0.016    | 0.001    |
| T cells regulatory (Tregs)   | 0.016    | 911.426       | 180.838           | 473.779 | 2.12E+04 | 619.741  | 0.005    | 0.000    |
| T cells gamma delta          | 0.000    | 0.000         | 0.006             | 0.000   | 0.000    | 1.79E+07 | 0.000    | 0.000    |
| NK cells resting             | 0.000    | 47.279        | 7.09E+06          | 0.148   | 0.004    | 0.160    | 0.169    | 5.22E+17 |
| NK cells activated           | 0.119    | 0.000         | 1.44E+06          | 0.004   | 11.271   | 24.971   | 0.159    | 0.000    |
| Monocytes                    | 0.130    | 0.000         | 0.165             | 0.000   | 0.020    | 0.008    | 2.640    | 0.000    |
| Macrophages M0               | 1.297    | 10.448        | 228.517           | 0.283   | 7.436    | 3.285    | 1.415    | 107.278  |
| Macrophages M1               | 0.000    | 0.021         | 0.273             | 148.932 | 3068.960 | 2.517    | 0.005    | 0.003    |
| Macrophages M2               | 17.506   | 5.241         | 0.095             | 11.871  | 3.672    | 4.362    | 3.703    | 185.137  |
| Dendritic cells resting      | 58.041   | 0.000         | 0.000             | 0.120   | 0.005    | 1.894    | 4.154    | 324.659  |
| Dendritic cells activated    | 2.57E+08 | 2.02E+09      | 2409.459          | 0.000   | 2.93E+18 | 1295.742 | 3905.087 | 0.000    |
| Mast cells resting           | 0.257    | 0.019         | 0.588             | 8.576   | 0.004    | 0.078    | 0.175    | 0.000    |
| Mast cells activated         | 34.007   | 6.309         | 0.157             | 16.627  | 0.057    | 2.814    | 137.645  | 9.24E+05 |

**Table 5.27: Cox regression p values for Hazard ratios in Table 5.26.**

|                              | GSE2603 | GSE6532<br>AB | GSE6532<br>plus 2 | GSE7390 | GSE11121 | GSE12276 | GSE20685 | GSE58812 |
|------------------------------|---------|---------------|-------------------|---------|----------|----------|----------|----------|
| Cell name                    | Cox p   | Cox p         | Cox p             | Cox p   | Cox p    | Cox p    | Cox p    | Cox p    |
| B cells naive                | 0.328   | 0.676         | 0.247             | 0.343   | 0.923    | 0.232    | 0.358    | 0.560    |
| B cells memory               | 0.327   | 0.451         | 0.664             | 0.182   | 0.901    | 0.646    | 0.053    | 0.128    |
| Plasma cells                 | 0.429   | 0.439         | 0.644             | 0.650   | 0.156    | 0.159    | 0.017    | 0.598    |
| T cells CD8                  | 0.484   | 0.651         | 0.938             | 0.927   | 0.972    | 0.565    | 0.039    | 0.410    |
| T cells CD4 naive            | 0.997   | 0.981         | 0.972             | 0.602   | 0.934    | 0.573    | 0.709    | 0.170    |
| T cells CD4 memory resting   | 0.347   | 0.158         | 0.616             | 0.340   | 0.964    | 0.214    | 0.366    | 0.147    |
| T cells CD4 memory activated | 0.997   | 0.155         | 0.176             | 0.688   | 0.577    | 0.284    | 0.127    | 0.037    |
| T cells follicular helper    | 0.838   | 0.162         | 0.823             | 0.129   | 0.280    | 0.136    | 0.232    | 0.155    |
| T cells regulatory (Tregs)   | 0.561   | 0.101         | 0.584             | 0.242   | 0.041    | 0.141    | 0.457    | 0.410    |
| T cells gamma delta          | 0.754   | 0.361         | 0.580             | 0.663   | 0.475    | 0.216    | 0.074    | 0.008    |
| NK cells resting             | 0.461   | 0.390         | 0.297             | 0.773   | 0.399    | 0.670    | 0.758    | 0.044    |
| NK cells activated           | 0.822   | 0.072         | 0.123             | 0.617   | 0.752    | 0.293    | 0.760    | 0.028    |
| Monocytes                    | 0.917   | 0.389         | 0.813             | 0.180   | 0.761    | 0.423    | 0.915    | 0.273    |
| Macrophages M0               | 0.891   | 0.083         | 0.022             | 0.527   | 0.286    | 0.152    | 0.749    | 0.001    |
| Macrophages M1               | 0.096   | 0.274         | 0.791             | 0.078   | 0.030    | 0.558    | 0.020    | 0.083    |
| Macrophages M2               | 0.357   | 0.370         | 0.311             | 0.251   | 0.661    | 0.157    | 0.369    | 0.040    |
| Dendritic cells resting      | 0.734   | 0.208         | 0.261             | 0.807   | 0.540    | 0.900    | 0.920    | 0.455    |
| Dendritic cells activated    | 0.109   | 0.083         | 0.399             | 0.315   | 0.165    | 0.079    | 0.043    | 0.172    |
| Mast cells resting           | 0.826   | 0.287         | 0.836             | 0.474   | 0.266    | 0.033    | 0.550    | 0.056    |
| Mast cells activated         | 0.663   | 0.539         | 0.677             | 0.609   | 0.607    | 0.791    | 0.145    | 0.010    |

Naive B cells showed good prognosis in 5 datasets and bad prognosis in 3 datasets. Memory B cells showed good prognosis in 6/8 datasets. Plasma cells showed good prognosis in 5 datasets and in 3 datasets, it showed bad prognosis with one being significant. CD8 T cells showed good prognosis in 5 datasets with one being significant. A similar pattern was seen for CD8 and CD4 naive T cells. CD4 memory resting cells showed bad prognosis in 6 datasets. CD4 memory activated, follicular helper gamma delta T cells showed good prognosis in 7 datasets. Regulatory T cells showed bad prognosis in 5 datasets with one being significant. Among macrophages, M0 and M2 showed bad

prognosis in a greater number of datasets (7/8 datasets) while M1 showed good prognosis in a higher number of datasets (5/8 datasets). Additionally, resting cells showed both good prognosis and bad prognosis but activated dendritic cells showed worse survival in 6 datasets with one being significant. Resting mast cells showed good prognosis (7/8 datasets) and activated mast cells showed worse survival (6/8 datasets). Eosinophils showed good prognosis overall (5/8 datasets) while in 3 datasets, HR could not be evaluated. Lastly, neutrophils showed worse survival for 6 datasets with 2 being significant.

In conclusion, we did not find CIBERSORT generated hematopoietic cell percentages of no prognostic importance significantly.

### **5.5. Prognostic importance of INHBA and MLANA in melanoma**

Survival analysis using CNCL showed that in the majority of datasets in breast cancer, CS/M phenotype showed worse prognosis when compared with NS/E phenotype. We wanted to check if a similar pattern will be found in another cancer as we have discussed the role of EMT and drug response in melanoma in our previous publications (Senses et al., 2017; Kucukkaraduman et al., 2020). Previously, in melanoma our studies have shown that INHBA and MLANA genes can classify, melanoma in invasive/proliferative, stem/non-stem and mesenchymal/epithelial phenotype similar to CNCL. To identify the prognostic role of INHBA and MLANA in melanoma, TCGA (Akbani et al., 2015) data was analyzed. TCGA data contained overall survival information for 458 patients along with AJCC pathologic T, N and M stage. We performed Cox regression analysis for each clinical parameter along with INHBA and MLANA (Table 5.28). INHBA showed good prognosis but insignificantly and MLANA showed bad prognosis with a borderline significance of 0.059. Other clinical parameters such as T, N and M stage showed bad prognosis significantly as expected.

**Table 5.28: Cox regression analysis for INHBA, MLANA and clinical parameters.**

| Parameter               | Sig.  | Exp(B) | 95.0% CI for Exp(B) |       |
|-------------------------|-------|--------|---------------------|-------|
|                         |       |        | Lower               | Upper |
| INHBA Continuous        | 0.285 | 0.966  | 0.907               | 1.029 |
| MLANA Continuous        | 0.059 | 1.038  | 0.999               | 1.078 |
| AJCC Pathologic M stage | 0.04  | 1.896  | 1.029               | 3.493 |
| AJCC Pathologic N stage | >0.00 | 1.366  | 1.197               | 1.559 |
| AJCC Pathologic T stage | >0.00 | 1.441  | 1.254               | 1.656 |

We then used LRMC analysis to divide patients in High and low expression groups using a cutoff. For INHBA, cutoff 0.137605 was selected and patients with higher expression than this were classified as “High” expression group and the rest of patients were classified as “Low” group. Upon Kaplan Meier analysis, High group showed better survival when compared with Low expressing group (Figure 5.33). For MLANA, cutoff 8.42132 was selected to classify patients into High and Low groups. The high group showed worse survival when compared with Low expressing group (Figure 5.34). Patient distribution, the median of survival and p value are shown in Table 5.29.

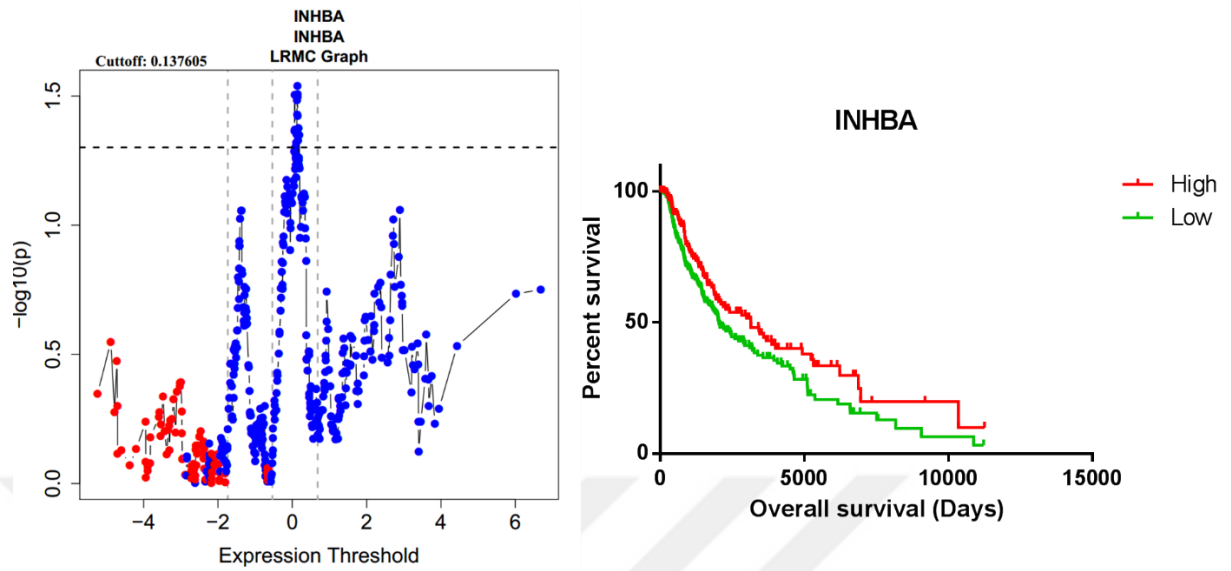


**Figure 5.33: Prognostic evaluation of INHBA in TCGA cohort.** Patients were classified into High and Low group using LRM analysis and High group showed better survival when compared with low expressing group.

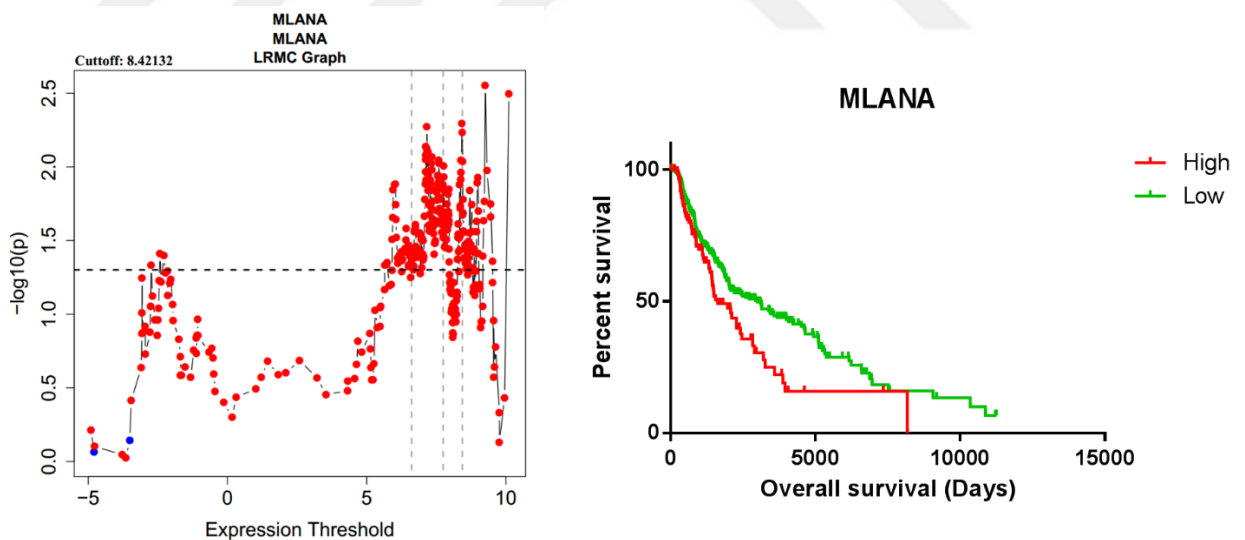


**Figure 5.34: Prognostic evaluation of MLANA in TCGA cohort.** Patients were classified into High and Low groups using LRM analysis and the High group showed worse survival when compared with Low group.

**Table 5.29: Patient distribution, median survival and Log rank p value for INHBA and MLANA classifications.**

|                 | INHBA  |      | MLANA |      |
|-----------------|--------|------|-------|------|
| Groups          | High   | Low  | High  | Low  |
| N               | 162    | 294  | 116   | 340  |
| Median Survival | 3139   | 2030 | 1618  | 2993 |
| Log Rank p      | 0.0289 |      | 0.005 |      |

Next we performed a multivariate analysis using backward Wald method, and patients' classification using INHBA was an independent survival predictor for TCGA patients along with clinical parameters (Table 5.30).

**Table 5.30: Multivariate analysis for TCGA with parameters significant with univariate analysis.**

| Backward<br>Wald | Parameters              | Sig.  | Exp(B) | 95.0% CI for<br>Exp(B) |       |
|------------------|-------------------------|-------|--------|------------------------|-------|
|                  |                         |       |        | Lower                  | Upper |
| Step 1           | AJCC Pathologic M stage | 0.073 | 2.081  | 0.935                  | 4.636 |
|                  | AJCC Pathologic N stage | >0.00 | 1.498  | 1.287                  | 1.742 |
|                  | AJCC Pathologic T stage | >0.00 | 1.501  | 1.302                  | 1.73  |
|                  | INHBA Categorical       | 0.045 | 0.717  | 0.518                  | 0.993 |
|                  | MLANA Categorical       | 0.106 | 1.326  | 0.942                  | 1.867 |
| Step 2           | AJCC Pathologic M stage | 0.084 | 2.026  | 0.91                   | 4.508 |
|                  | AJCC Pathologic N stage | >0.00 | 1.491  | 1.281                  | 1.735 |
|                  | AJCC Pathologic T stage | >0.00 | 1.511  | 1.312                  | 1.74  |
|                  | INHBA Categorical       | 0.025 | 0.691  | 0.501                  | 0.954 |

This analysis showed us that INHBA is a good prognosis marker of melanoma patients and MLANA is a bad prognosis marker. Additionally, patients' stratification using MLANA is an independent prognostic factor of clinical parameters.

Next we analyzed dataset GSE65904 (Cirenajwis et al., 2015) which contains both distant metastasis and disease specific survival information for 214 patients. We analyzed this dataset for prognostic importance for both MLANA and INHBA. Distant metastasis free survival data (DMFS) was available for 150 patients. Cox regression analysis showed that INHBA is good and MLANA is bad prognostic biomarker for melanoma distant metastasis free survival insignificantly (Table 5.31).

**Table 5.31: Cox regression analysis for INHBA and MLANA for GSE65904 DMFS**

| Parameter          | Sig.  | Exp(B) | 95.0% CI for Exp(B) |       |
|--------------------|-------|--------|---------------------|-------|
|                    |       |        | Lower               | Upper |
| INHBA/ILMN_1719547 | 0.689 | 0.646  | 0.076               | 5.474 |
| MLANA/ILMN_1710217 | 0.55  | 1.036  | 0.923               | 1.162 |

We then used LRMC analysis to determine the cutoffs for INHBA and MLANA. For INHBA cutoff 6.745493 was used and for MLANA, cutoff 12.64553 was used to divide patients in High and Low expression groups. INHBA showed good prognosis (Figure 5.35) and MLANA (Figure 5.36) showed bad prognosis upon Kaplan Meier analysis. Log rank p values for Kaplan Meier analysis are shown in Table 5.32.

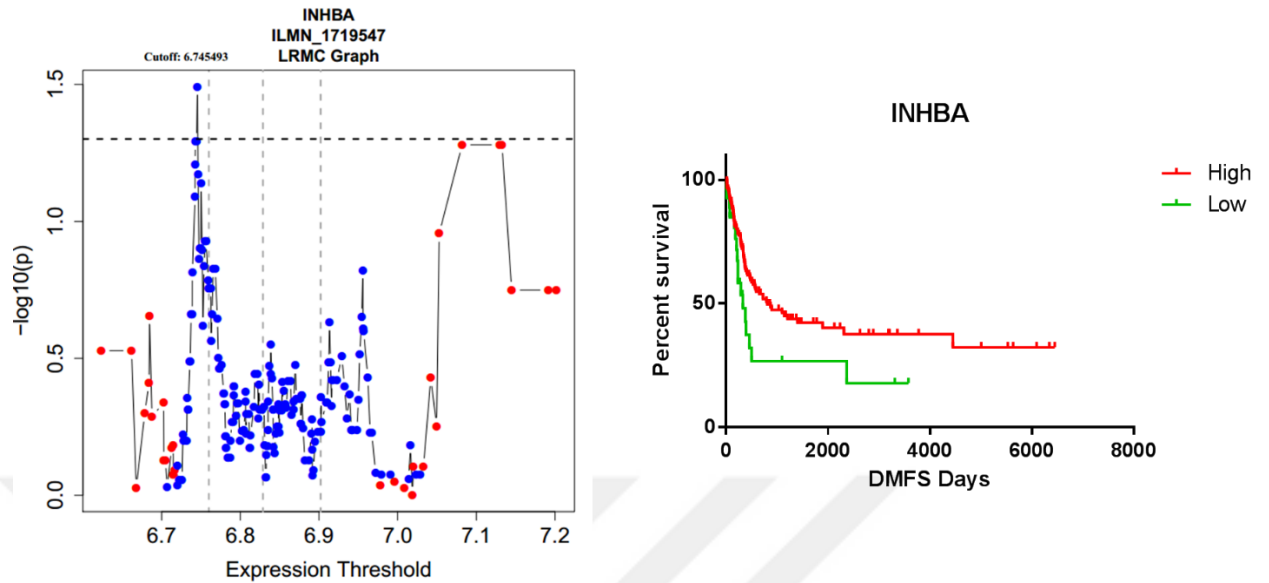


**Figure 5.35: Prognostic evaluation of INHBA in GSE65904 cohort for distant metastasis free survival data.** Patients were classified into High and Low groups using LRM analysis and High group showed better survival when compared with Low expressing group.

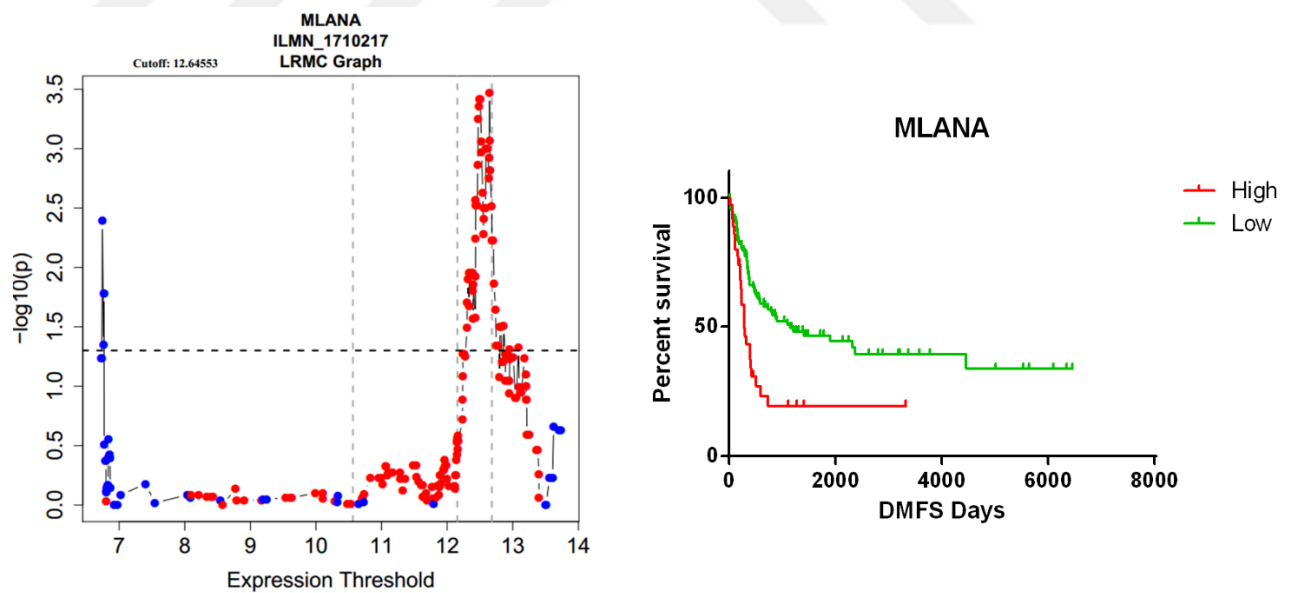


**Figure 5.36: Prognostic evaluation of MLANA in GSE65904 cohort for distant metastasis free survival data.** Patients were classified into High and Low groups using LRM analysis and High group showed worse survival when compared with Low expressing group.

**Table 5.32: Patient distribution, median survival and Log rank p value for INHBA and MLANA classifications in GSE65294 DMFS.**

| DMFS            | INHBA |     | MLANA  |      |
|-----------------|-------|-----|--------|------|
| Groups          | High  | Low | High   | Low  |
| N               | 123   | 27  | 36     | 114  |
| Median Survival | 811   | 331 | 288    | 1143 |
| Log Rank p      | 0.032 |     | >0.000 |      |

Additionally, GSE65294 dataset contained disease specific survival data as well. We used this to evaluate prognostic importance for INHBA and MLANA as well. Upon cox regression analysis, INHBA showed good survival and MLANA showed bad survival but p value was insignificant (Table 5.33).

**Table 5.33: Cox regression analysis for INHBA and MLANA for GSE65904 DSS**

| Parameter          | Sig.  | Exp(B) | 95.0% CI for Exp(B) |       |
|--------------------|-------|--------|---------------------|-------|
|                    |       |        | Lower               | Upper |
| INHBA/ILMN_1719547 | 0.412 | 0.437  | 0.061               | 3.15  |
| MLANA/ILMN_1710217 | 0.808 | 1.012  | 0.916               | 1.119 |

We then used LRMC analysis to determine the cutoffs for INHBA and MLANA for classifying patients into High and Low group separately. For INHBA cutoff 6.954407 (Figure 5.37) and for MLANA cutoff 12.807075 (Figure 5.38) was used to divide patients into High and Low groups. Upon Kaplan Meier analysis, High group for INHBA showed good prognosis when compared with Low group and for INHBA, opposite pattern was observed. Log rank p values are shown in Table 5.34.

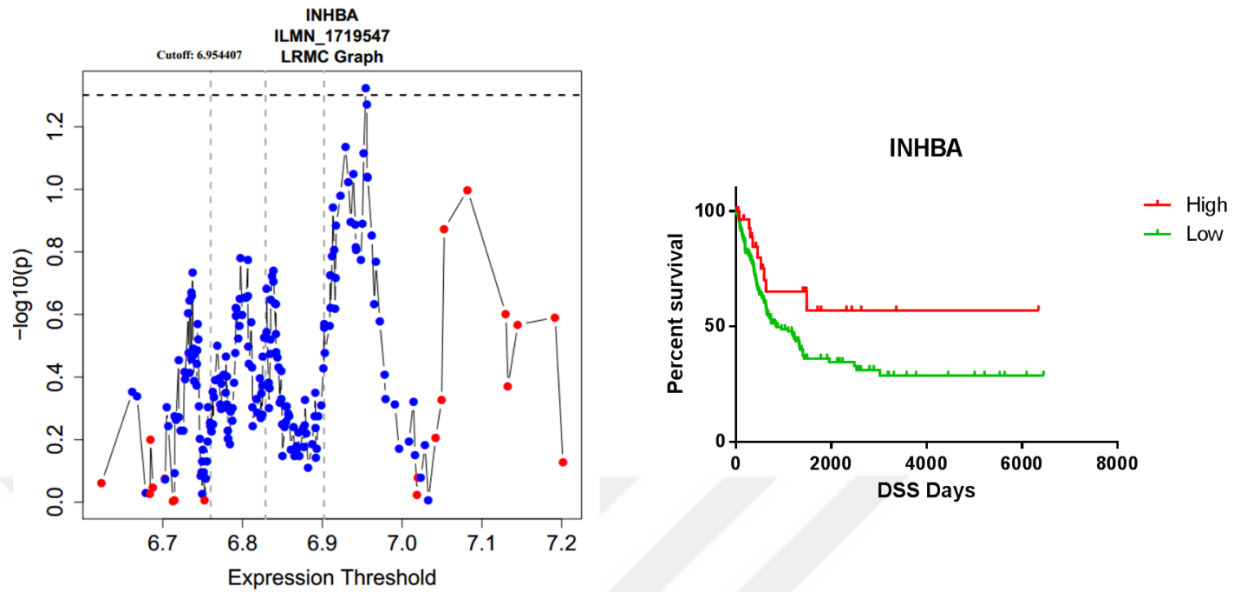


**Figure 5.37: Prognostic evaluation of INHBA in GSE65904 cohort for disease specific survival data.** Patients were classified into High and Low groups using LRM analysis and High group showed better survival when compared with Low expressing group.

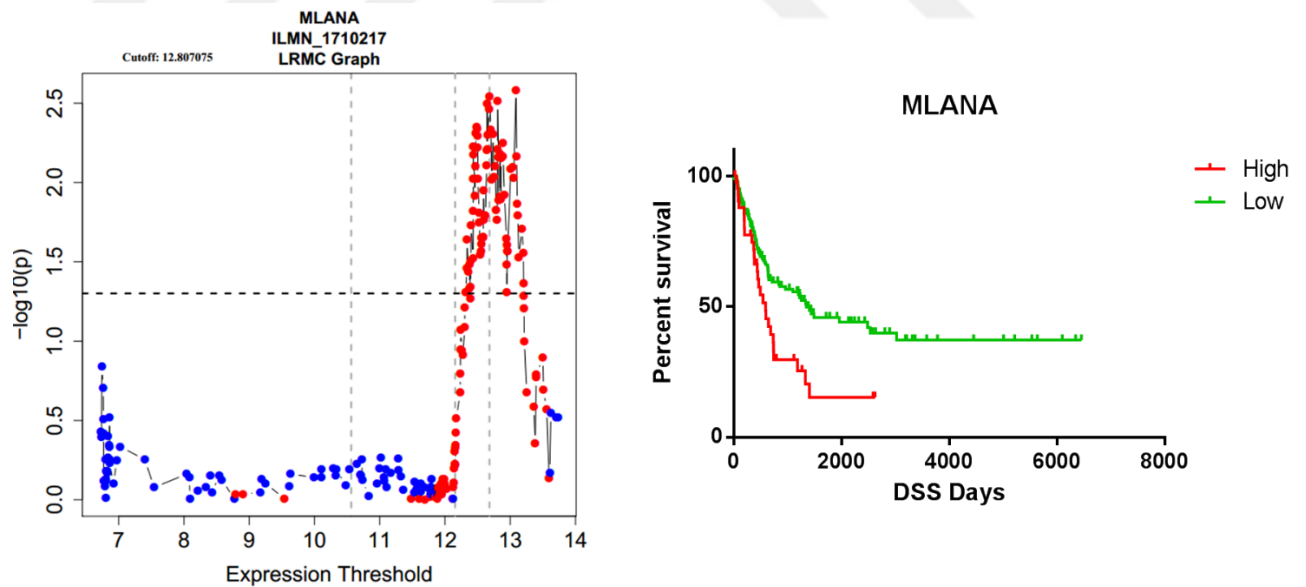


**Figure 5.38: Prognostic evaluation of MLANA in GSE65904 cohort for disease specific survival data.** Patients were classified into High and Low groups using LRM analysis and High group showed worse survival when compared with Low expressing group.

**Table 5.34: Patient distribution, median survival and Log rank p value for INHBA and MLANA classifications in GSE65904 DSS.**

| DSS             | INHBA |     | MLANA |      |
|-----------------|-------|-----|-------|------|
| Groups          | High  | Low | High  | Low  |
| N               | 30    | 180 | 43    | 67   |
| Median Survival | NA    | 849 | 591   | 1344 |
| Log Rank p      | 0.047 |     | 0.003 |      |

Nex, we validated the prognostic importance of INHBA and MLANA in dataset GSE59455 (Budden et al., 2016). This dataset contains expression data for 141 tumors with overall survival clinical data. All tissues were formalin fixed paraffin embedded and were diagnosed as stage 2 or greater. We performed Cox regression analysis for MLANA, BRAF mutation, NRAS mutation and Breslow depth (Table 5.35). Only Breslow depth showed significance in this analysis.

**Table 5.35: Cox regression analysis for INHBA, MLANA, BRAF mutation, NRAS mutation, and Breslow depth in dataset GSE59455.**

| Parameter             | Sig.  | Exp(B) | 95.0% CI for Exp(B) |       |
|-----------------------|-------|--------|---------------------|-------|
|                       |       |        | Lower               | Upper |
| INHBA/ILMN_1719547    | 0.38  | 0.91   | 0.739               | 1.122 |
| MLANA/ILMN_1710217    | 0.492 | 1.023  | 0.959               | 1.091 |
| BRAF and NRAS mutated | 0.137 | 0.872  | 0.728               | 1.045 |
| BRAF Mutation         | 0.86  | 1.019  | 0.83                | 1.251 |
| NRAS Mutation         | 0.074 | 0.82   | 0.659               | 1.02  |
| Breslow Depth         | 0.003 | 1.061  | 1.021               | 1.104 |

We next performed LRMC analysis for determining cutoffs for INHBA and MLANA to classify patients in High and low groups. Any cutoff for INHBA did not show any significance. And MLANA showed significance. For MLANA 12.5554474 was selected to divide patients into high and low groups (Figure 5.39) and Kaplan Meier analysis showed that High group showed better survival when compared with Low group as expected. Log rank p value is shown in Table 5.36.

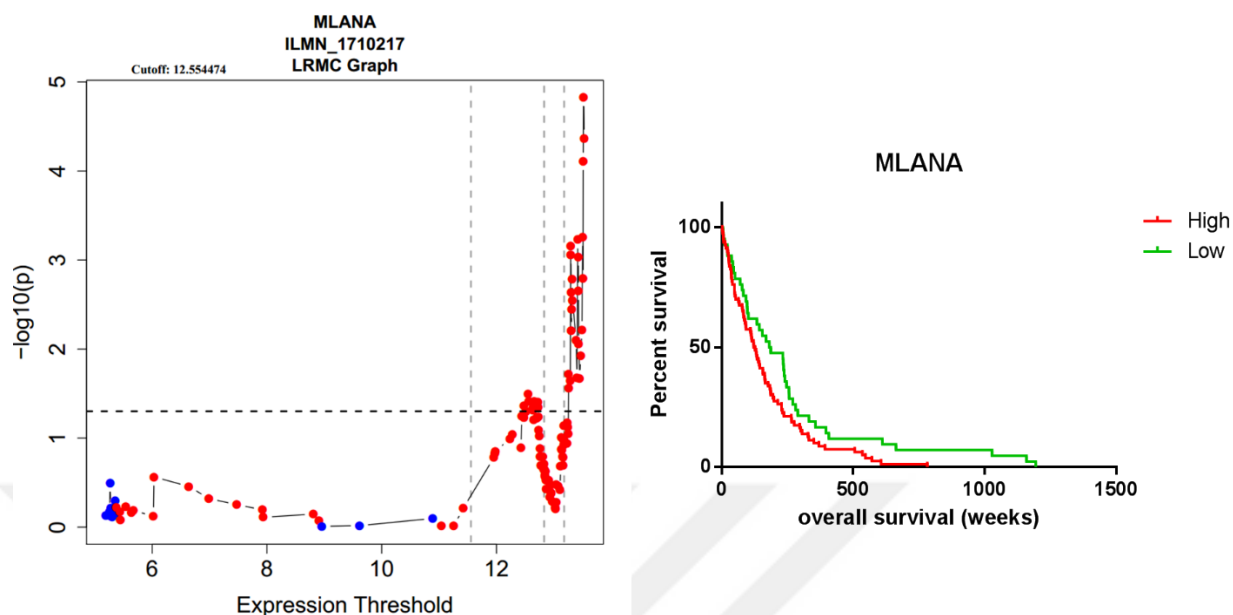


**Figure 5.39: Prognostic evaluation of MLANA in GSE59455 cohort for overall survival data.** Patients were classified into High and Low groups using LRM analysis and High group showed worse survival when compared with Low expressing group.

**Table 5.36: Patient distribution, median survival and Log rank p value for MLANA classifications in GSE59455 OS.**

| OS              | MLANA   |     |
|-----------------|---------|-----|
| Groups          | High    | Low |
| n               | 80      | 42  |
| Median Survival | 122.143 | 181 |
| Log Rank p      | 0.032   |     |

Next, we performed multivariate analysis for MLANA classification and Berslow depth. Multivariate analysis showed that both MLANA and Berslow depth are independent prognostic factors for melanoma overall survival (Table 5.37).

**Table 5.37: Multivariate analysis for MLANA classification and Berslow depth in dataset GSE59455.**

| Variables            | Sig. | Exp(B) | 95.0% CI for Exp(B) |       |
|----------------------|------|--------|---------------------|-------|
|                      |      |        | Lower               | Upper |
| MLANA\classification | 0.04 | 1.315  | 1.013               | 1.708 |
| Breslow Depth        | 0.01 | 1.054  | 1.013               | 1.097 |

Next, we analyzed dataset GSE54467 (Jayawardana et al., 2015). This dataset expression data for 84 patients and samples were collected from lymph node specimens with macroscopic tumor. This data contained overall survival data. INHBA expression was missing from the normalized file downloaded from GEO so we performed the analysis only with MLANA. First, cox regression analysis was performed and MLANA showed bad prognosis insignificantly (Table 5.38) as expected.

**Table 5.38: Cox regression analysis for MLANA in GSE54467**

| Parameter          | Sig.  | Exp(B) | 95.0% CI for Exp(B) |       |
|--------------------|-------|--------|---------------------|-------|
|                    |       |        | Lower               | Upper |
| MLANA/ILMN_1710217 | 0.571 | 1.045  | 0.898               | 1.215 |

We next performed LRMC analysis and chose the cutoff 13.29367 to divide patients in High and Low group for MLANA (Figure 5.40). Kaplan Meier analysis showed that high expression group showed worse survival when compared with Low expression group. Log rank p value is shown in Table 5.39.

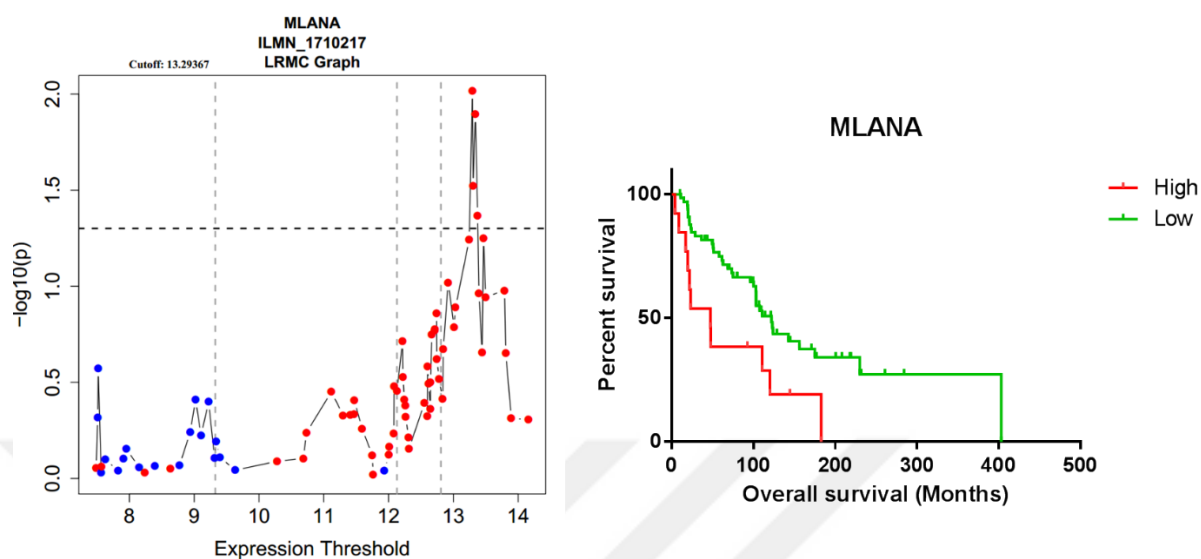


**Figure 5.40: Prognostic evaluation of MLANA in GSE54467 cohort for overall survival data.** Patients were classified into High and Low groups using LPMC analysis and High group showed worse survival when compared with Low expressing group.

**Table 5.39: Patient distribution, median survival and Log rank p value for MLANA classifications in GSE54467 OS.**

| OS              | MLANA |        |
|-----------------|-------|--------|
| Groups          | High  | Low    |
| n               | 13    | 66     |
| Median Survival | 47.28 | 122.17 |
| Log Rank p      | 0.01  |        |

In conclusion, this analysis showed that high expression of INHBA and MLANA is associated with increased and decreased survival of melanoma patients and this observation was validated in several cohorts. Therefore, these findings in melanoma supported our results in breast cancer.

## 6. Discussion

### 6.1. CNCL based classification of breast cancer into stemness and EMT subtypes

Previously, in literature, a lot of markers have been reported to identify breast cancer subtypes such as stem and non-stem along with PAM50 molecular subtypes (Parker et al., 2009; Bai et al., 2018). In our analysis, CNCL showed a high correlation with breast cancer stem markers, CD44 and CD24, but not with ALDH. EMT markers-based classification has been shown to exist across several cancers including breast cancer (Tan et al., 2014). Additionally, CNCL showed a great correlation with a previously published gene list (Tan et al., 2014) and it can classify both cell lines and tumor tissue.

Cellular differentiation has been shown to play a strong role in inducing sensitivity/resistance to therapy (Housman et al., 2014). However, breast tumors have an increased tendency to switch phenotype, referred to as phenotype plasticity. This tendency is attributed to both de-differentiation and trans-differentiation states (Gupta et al., 2019). We showed in our analysis, both *in silico* and *in vitro*, that phenotype plasticity occurs both in cell lines and primary tumors without an exception and this switch leads to change in chemotherapy response. Both ZEB1 and SNAIL1 are mesenchymal markers and their induction in cells is related to EMT and chemotherapy resistance (Shang et al., 2013). Additionally, ZEB1 has been associated with tumor cell plasticity (Krebs et al., 2017). Based upon these observations, downregulating these genes results in increased sensitivity to chemotherapy as observed in our *in silico* analysis and knockdown experiments *in vitro*.

Metastasis plays a major role in cancer-related mortality. Although, EMT plasticity has been related to metastasis but recent data shows that EMT plasticity of a tumor is not related to its metastatic potential (Pastushenko et al., 2018). Instead, EMT plasticity rate is a better indicator of drug responses (Gupta et al., 2019). It is, therefore, important to quantify EMT transition rate to predict chemotherapeutic response and CNCL is well-suited candidate in this regard for validation in bigger cohorts.

Gene signatures in breast cancer have been previously associated with drug response prediction (Hsu et al., 2013) and EMT has been associated with resistance to EGFR

inhibitor (gefitinib) (Weng et al., 2019). Based on our analysis, we suggest a combinational therapy of two drugs targeting NS/E cells and CS/M cells separately. We know in line with our previous analysis (Isbilen et al., 2013) that Midostaurin and Lapatinib are candidates for such combinational therapy. Although EMT is a shared signature among many tumor types, our results for breast cancer analysis were different from our previous analysis for melanoma and colorectal cancer (Demirkol et al., 2017; Senses et al., 2017). This discrepancy might be explained as CS/M and NS/E phenotypes defined here are different from other tumors as these phenotypes exist as a hybrid state in between epithelial and mesenchymal states. Indeed most of the CS/M cells of the colon are similar to NS/E cells of breast cancer and Melanoma cells show a similar phenotype to CS/M phenotype (Tan et al., 2014).

CNCL clearly showed that NS/E phenotype cells pre-therapy, switched to CS/M phenotype post-therapy both *in silico* and *in vitro*. This observation elucidates the reason for secondary treatment resistance in breast cancer. Previously post neo-adjuvant chemotherapy, cells with higher proliferative potential are related with sensitivity to chemotherapy (Klintman et al., 2016). Moreover, post-chemotherapy breast cancer cells show higher stemness when compare with pre-therapy samples (Liu et al., 2018b). These observations are in line with our survival analysis, as post neo-adjuvant Taxane based chemotherapy, CS/M phenotype showed worse survival when compared to NS/E phenotype. In light of these results, it is important to further investigate and determine if neo-adjuvant taxane treatment switches cells from NS/E phenotype to CS/M phenotype, increasing drug resistance in treated patients and decreasing survival of these patients. Additionally, our gene list can be used to differentiate among NS/E and CS/M phenotypes providing the opportunity in selecting a suitable treatment for each phenotype.

To summarize our findings, our identified list CNCL gene list can distinguish among different phenotypes of breast cancer induced through cancer plasticity including stemness and EMT phenotypes. This phenotypic switch correlates with cytotoxicity profiles of breast cancer and we propose that epithelial and non-stem phenotypes should be treated with Lapatinib, moreover mesenchymal and stem phenotypes should be treated with Midostaurin. CNCL is a unique signature and it was validated both *in silico* and *in vitro*.

Lastly, CNCL does not show any prognostic importance but its association with the prognosis of patients treated with neo adjuvant taxane therapy, signifies its importance and is worth validation in future studies *ex vivo*.

## **6.2. Breast cancer plasticity**

Apart from stemness and EMT plasticity, breast cancer also shows plasticity in terms of clinical and molecular classification. Among IHC clinical classification, breast tumors switch their status regarding ER, PR and HER2 biomarkers. IHC status is so unstable that almost 33% of breast cancer patients switch their IHC status considering all 3 markers in both directions (both positive to negative and negative to positive switching). Additionally One out of seven HER2<sup>+</sup> patients shows a switch in HER2 status(Lindstrom et al., 2012). Similar switching is shown by molecular classification achieved through PAM50 subtyping. This switching can result from the difference from two different biopsy methods. As we showed in dataset GSE32518, 22% of samples (8/37) showed a difference in molecular classification between matched cancer specimens collected through fine needle aspiration and core needle biopsies. Core needle biopsies sample are rich in stroma when compared with fine needle aspiration biopsies. Bianchini et al. showed that although if the biopsies are from the same tumor but different collection techniques are used and when corresponding samples were compared for their gene expression, immune related metagenes showed prognostic significance, suggesting selection of the biopsy method can provide additional information in selecting treatment plan for breast cancer patients as immune response can give useful prognostic information. Among the analyzed samples, Basal was the most stable subtype while Normal subtype showed most switching (Bianchini et al., 2010).

Among other factors that can lead to breast cancer plasticity, the most important is drug treatment (similar to our results with CNCL analysis). These observations are validated in 3 different datasets in which corresponding tumor samples were available both before and after therapy. In dataset GSE10281, gene expression data of 18 patients is available and among these patients, 39% of samples changed their molecular subtype. Creighton et al. reported that switched samples converted to a more mesenchymal phenotype similar to our

results shown in CNCL project (Creighton et al., 2009). In dataset GSE28844, 28 patients' expression data is available and these patients were treated with anthracyclines and taxanes in neoadjuvant setting. Here 36% of samples switched their molecular class. Contrary to our other analyzed datasets, most samples here switched to normal like phenotype. Vera-Ramirez et al. showed that genes related to extra-cellular matrix, oxidative stress response and developmental processes. These pathways are often related with drug response and EMT (Vera-Ramirez et al., 2013). Dataset GSE32072 contained 4 control samples which were not treated with any neoadjuvant therapy. As expected, none of these samples changed their molecular subtype throughout disease progression. Additionally, 9 patients were treated with anthracycline based therapy, 8 were treated with anthracycline/taxane based therapy and 2 were treated with trastuzumab therapy. Normal was the most unstable class and basal was the most stable. Gonzalez-Angulo et al showed that when pre and post treatment differentially expressed genes were related to Notch and Wnt signaling (Gonzalez-Angulo et al., 2012). These results show that Normal subtype is more prone to switch when compared with other subtypes and Basal subtype is most stable one. Additionally, residual cancer tissues show more mesenchymal features or the upregulated pathways are related to these. Based upon observations regarding subtype switching and acquisition of mesenchymal features in residual cancer tissue, one can conclude that both of these phenomena might play a role for each other and are related.

Moreover, tumor cells upon metastasis, also show molecular subtype switching. We analyzed 2 datasets in which both primary and metastatic tissue gene expression data was available. GSE125989 contained 16 patients' gene expression data for both primary and metastatic tissue. 50% of these samples switched molecular class upon metastasis. Normal subtype was the most unstable subtype. Similarly, dataset GSE110590 contains expression data for 16 patients with multiple metastases. Among these samples, Basal was the most stable subtype as most of the metastases tissue was also subtyped as Basal as well. And LumA was the most unstable subtype. Similar results were obtained by Cejalvo et al. authors compared 123 primary tissues with their respective metastasis and found that Basal was the most stable phenotype in terms of molecular switching upon metastasis, followed by HER2, then LumB and lastly LumA (Cejalvo et al., 2017).

So here we propose that patients treated with neo adjuvant therapy, should be re-subtyped for molecular classification to better identify and suggest appropriate treatment strategy. Similar is the case for metastatic events, so while treating metastasis, metastasis molecular subtype should always be considered in treatment strategy.

### **6.3. Chemosensitivity markers of paclitaxel, cisplatin and doxorubicin in breast cancer**

Our CNCL gene list can differentiate between CS/M and NS/E phenotype. So, we next we asked if there is a cytotoxicity difference between TNBC CSC phenotype and Non-CSC phenotype for paclitaxel, cisplatin and doxorubicin which are commonly used for TNBC treatment. For this analysis, we used 7 TNBC Non-CSC and 4 TNBC CSC cell lines. Upon drug screening, we could not find any significant cytotoxicity difference between these two groups. Similar results were obtained from CCLE and CGP datasets for docetaxel (paclitaxel), cisplatin and doxorubicin. Although previously we showed that CNCL gene list can clearly differentiate between drug resistant and sensitive cells but this discrepancy can be explained using innate resistance and acquired resistance. We believe that treatment of cells with these drugs over a prolonged period can shift their phenotype from CSC to Non-CSC/epithelial to mesenchymal as shown previously for doxorubicin (Saxena et al., 2011), paclitaxel (Gupta et al., 2009a) and cisplatin (Zeng et al., 2020). And our CNCL does not identify innate resistance to these drugs but rather identifies acquired resistance over a prolonged period. Then we aimed to identify genes that could be used as biomarkers for these drug sensitivity. And using these markers linear regression analysis models were generated which could predict IC50 for cell lines using gene expression values of those markers. Previously in several studies, Cox regression analysis has been used to calculate the risk of relapse/prognosis (Parker et al., 2009; Klein et al., 2013; Li et al., 2016). Popiołkiewicz et al. used a similar regression analysis to correlate IC50 *in vitro* data with LD50 *in vivo* for anticancer drugs (Popiołkiewicz et al., 2005). Brown et al used similar analysis to generate regression model for Adriamycin inhibitors (Brown et al., 1995).

### 6.3.1. Paclitaxel biomarkers

We selected genes that showed high correlation values with *in vitro* cytotoxicity data. These genes were used to generate 17 linear regression models and only three of these were further validated. The first validation model has two genes (AMD1 and SLC25A28). AMD1 showed a positive correlation (sensitive marker) and SLC25A28 showed a negative correlation (resistant marker) with paclitaxel activity area. Heng et al. reported AMD1 to be upregulated in breast cancer which showed pathological characteristics of necrosis, inflammation, nuclear polymorphism and high mitotic count in the same cancer (Heng et al., 2017). The role of AMD1 is not well studied in breast cancer but in melanoma, AMD1 silencing resulted in increased EMT and metastasis (Zingg et al., 2015). So, an increase in AMD1 expression in breast cancer tumors with higher mitotic count along with its tumor suppressor properties, makes it an ideal biomarker for paclitaxel which targets microtubule leading to defects in the cell cycle. SLC25A28 is responsible for transporting iron to inner mitochondria and previously was reported as a sensitivity inducer for photodynamic therapy (Jong and McKeage, 2014). SLC25A28 is a member of solute carrier family 25. In cancer, solute carriers play a dual role as they transport both anti-cancer drugs and nutrients to cancer cells and thus their role in sensitivity/resistance to the drug is drug dependent (Li and Shu, 2014). In pancreatic cancer, Kang et al. reported that degradation of SLC25A28 results in decreased growth of cancer cells (Kang et al., 2019). In our analysis, we found SLC25A28 as a paclitaxel resistance marker.

Second model consisted of TSPO and WDR61. WDR61 showed a positive correlation (sensitive marker) and TSPO showed a negative correlation (resistant marker) with paclitaxel activity area. WDR61 is a member of PAF1 complex and plays an important role in transcription, DNA repair and cell cycle. It was reported as overexpressed in breast cancer at the protein level when tumors were compared with normal adjacent tissue (Da Costa et al., 2015). TSPO protein is a part of the mitochondrial permeability pore complex and plays an important part in apoptosis regulation. Its expression is increased in cancer when compared with normal tissue and also is highly expressed in metastasis in several cancers including breast cancer (Austin et al., 2013).

The third model consisted of LHX2 and WDR61. WDR61 is common in models two and three. LHX2 showed a positive correlation (sensitive marker) with the paclitaxel activity area. LHX2 is involved in cell differentiation, proliferation and embryogenesis (Shi et al., 2015). Increased expression of LHX2 is increased during cancer development and it plays an important role in metastasis in breast cancer (Kuzmanov et al., 2014). In non-small cell lung cancer, LHX2 is targeted by miR-1238 which leads to decreased cell proliferation (Shi et al., 2015). It is important to note here that selected genes which were paclitaxel sensitivity markers show involvement in proliferation and resistant genes are involved in mitochondrial mechanisms.

### **6.3.2. Cisplatin Biomarkers**

We selected 16 genes that showed a high correlation with cisplatin cytotoxicity data and used these genes to generate 9 linear regression models. Only 2 of these models were validated consistently. The first model consisted of two genes which are C14orf169 (RIOX1, MAPJD, HOXD9) and DOHH. Both genes were positively correlated (resistance markers) with cisplatin IC50. C14orf169 also known as RIOX1 is upregulated in non-small cell lung cancer and Suzuki et al. showed that this is a member of MYC complex and transactivates kinases and cell signal transducers which are related with cancer cells proliferation (Suzuki et al., 2007). RIOX1 also shows selective expression in colorectal cancer and its high expression shows increased metastasis and invasion. Additionally, it was reported as an independent prognostic factor for colorectal cancer (Nishizawa et al., 2017). RIOX1 is upregulated in lung cancer when compared with normal tissue and patients with lower expression of RIOX1 show a better survival. Additionally, authors reported RIOX1 as a potential target for treating platinum resistant tumors (Pernía et al., 2020). Role of RIOX1 in breast cancer is clear but literature in lung cancer is in agreement with our finding that this gene is related with cisplatin resistance. DOHH is involved in the cell cycle and previously ciclopirox is reported as its inhibitor (Mémin et al., 2014). Using a knockout mice model, Sievert et al. showed that DOHH plays an important role in embryonic development and malignant transformation (Sievert et al., 2014). DOHH knock out mice do not survive which shows the high importance of DOHH in embryogenesis (Pällmann et al., 2015). Epis et al. showed that increased levels of miR-331-3p and miR-

642-5p decrease DOHH expression leading to decreased proliferation in prostate cancer (Epis et al., 2012). Liu et al. showed the blocking DOHH in cervical squamous carcinoma leads to apoptosis of carcinoma cells (Liu et al., 2018c). DOHH role in breast cancer and cisplatin sensitivity is not well explored.

Second model also consisted of 2 genes C14orf169 (common with the first model) and RBBP4. RBBP4 showed a negative correlation (sensitive marker) with cisplatin IC50. RBB4 regulates gene transcription, cell cycle and proliferation. In breast cancer, increased expression of RBBP4 was related with decreased overall survival and increased lymph node metastasis (Guo et al., 2020). Similar results were reported for hepatic metastasis and colon cancer prognosis (Li et al., 2019). In our results increase in RBBP4 expression was related to cisplatin sensitivity.

### **6.3.3. Doxorubicin Biomarkers**

We selected 19 genes which showed a significant correlation with doxorubicin IC50 data and used these to generate 10 models among which two models were selected. The first model consisted of 2 genes (FAM136A//LOC100287852, NME5). Both genes were negatively correlated (sensitive biomarkers) with doxorubicin IC50. Zhao et al. showed that FAM136A knockdown resulted in decreased proliferation and migration along with increased apoptosis. Additionally, FAM136A was proposed as an independent prognostic factor for lymph node metastasis (Zhao et al., 2020). NME5 is upregulated in breast cancer stem cells upon irradiation (Konge et al., 2018). Bensimon et al. reported downregulation of NME5 in non-stem cells upon irradiation (Bensimon et al., 2016). In TNBC, high expression of NME5 was related with poor prognosis (Zhai et al., 2019). In our analysis, both genes were sensitivity markers for doxorubicin treatment.

The second model consisted of SCGB1D2 and FZD3. Both genes were negatively correlated (sensitive biomarkers) with doxorubicin IC50. SCGB1D2 gene is specifically expressed in breast cancer (Watson et al., 1998). Smid et al. found SCGB1D2 commonly upregulated in HER2 patients and patients with bone relapse in breast cancer when compared with patients with other breast cancer subtypes and patients with site of relapse (Smid et al., 2008). SCGB1D2 shows a high correlation with metastatic marker SCGB2A2

and both proteins form heteroduplex. Additionally Carter et al. proposed that serum antibodies against SCGB1D2 can be used as breast cancer diagnostic marker (Carter et al., 2003). Doebar et al. reported that SCGB1D2 was upregulated in patients with ductal carcinoma *in situ* (DCIS) tumors with invasive breast cancer when compared with DCIS tumors only (Doebar et al., 2017). The upregulation of SCGB1D2 was associated with doxorubicin sensitivity in our data. FZD3 is mapped to chromosome 8p21 and this region plays an important role in Wnt signaling. Mo et al. reported that the downregulation of FZD3 resulted in Wnt/ $\beta$ -Catenin signaling inhibition along with decreased proliferation and migration in breast cancer cells (Mo et al., 2019). Additionally, Bell et al. proposed FZD3 as a breast cancer onset marker and its expression was higher in metastasis when compared with primary tumor tissues (Bell et al., 2017). Szwed et al. showed that leukemia cells upon treatment with doxorubicin-transferrin conjugate, showed decreased expression of FZD3 (Szwed et al., 2015). In our analysis, both genes were sensitivity markers for doxorubicin treatment.

In conclusion, identified models for doxorubicin, paclitaxel and cisplatin have the potential for predicting IC50s for each drug and will be validated in future studies in further experiments both *in silico* and *in vitro*.

#### **6.4. Biomarkers for metastasis**

We identified 20 genes as metastasis biomarkers in breast cancer. 10 genes were termed as good prognostic genes as their high expression was found related with decreased metastasis and the other 10 were termed as bad prognostic genes as their higher expression was related to increased metastasis.

##### **6.4.1. Good prognostic genes**

Epoxide hydrolase 2 (EPHX2) was identified as a good prognosis gene as its high expression was found to be related to decreased metastasis. Biologically EPHX2 plays its role in epoxy fatty acid metabolism and is involved in cardiovascular and renal systems along with pain and inflammation response (Tanaka et al., 2008). In breast cancer in a previous study by Ellsworth et al., authors showed that EPHX2 is upregulated upon

metastatic suppression and for LN metastasis, its expression was higher in tumors with lymph node negative samples when compared with lymph node positive samples (Ellsworth et al., 2010). Similarly, Thomassen et al. showed in their meta-analysis that EPHX2 is related with drug metabolism, response to inflammation and reactive oxygen species metabolism and acts in metastasis suppression capacity (Thomassen et al., 2009). Increased breast cancer proliferation and metastasis are associated with high glucose levels (Sun et al., 2019) and Oguro et al. showed that EPHX2 is downregulated in response to increased glucose levels under the influence of specificity protein 1 (sp1) (Oguro et al., 2015). In addition to these findings, EPHX2 plays an important role in breast cancer recurrence as shown by Naoi et al. (Naoi et al., 2011).

Thymocyte nuclear protein 1 (THYN1, THY28) is an important regulator of biological processes as in B cell development and is found conserved across species from bacteria to mammals (Fujita et al., 2015). In Ramos B lymphoma, downregulation of THYN1 is associated with apoptosis (Jiang et al., 2003). In our analysis, THYN1 was associated with decreased metastasis but its role in breast cancer is not fully explored. Loi et al. associated its role in breast cancer with better survival (Loi et al., 2009). Naoi et al. showed a similar association with decreased breast cancer recurrence and THYN1 expression (Naoi et al., 2011). Roschke et al. showed a negative correlation between THYN1 expression and numerical chromosomal heterogeneity in cancer cell lines (NCI60 dataset). Additionally, authors found increased chromosomal instability leads to EMT, invasiveness and metastasis (Roschke et al., 2008).

ADP ribosylation factor like GTPase 6 interacting protein 5 (ARL6IP5, JWA) is microtubule associated protein and plays an important part in biological processes such as stress response, differentiation and apoptosis. Importantly, ER mediates ARL6IP5 expression (Chen et al., 2005). Chen et al. showed that an increase in H<sub>2</sub>O<sub>2</sub> levels increased ARL6IP5 expression in NIH3T3 and HELF cell lines (Chen et al., 2007). In breast cancer cell line MDA-MB-231 cell line, upon ARL6IP5 knockdown, cells showed increased invasion, migration and decreased apoptosis (Chen et al., 2015). Similarly, an important breast cancer marker's expression (CXCR4) was inversely related to ARL6IP5. And an increased expression of ARL6IP5 showed decreased migration and invasion due to CXCR4

proteomic degradation (Xu et al., 2018b). Similar to these results in breast cancer, ARL6IP5 plays an important part in decreasing metastasis and invasion in melanoma and gastric cancer (Bai et al., 2010; Chen et al., 2013c; Lu et al., 2013b).

CAMP responsive element binding protein like 2 (CREBL2) is an important metabolic regulator (Tiebe et al., 2019). Boersma et al. compared two cell lines with low (T47D) and high (MDA-MB-435) metastatic potential and CREBL2 expression was significantly downregulated in a high metastatic cell line (Mellick et al., 2002). In inflammatory breast cancer which is a highly aggressive subtype of breast cancer CREBL2 expression is found to be downregulated (Boersma et al., 2008). Similarly, Cava et al. showed decreased expression of CREBL2 in grade three breast cancer samples (Cava et al., 2014). Importantly Myhre et al. also identified CREBL2 downregulation in metastatic patients when compared to non-metastatic patients (Myhre et al., 2010). CREBL2 also plays an important role in drug resistance as it was found downregulated OVCAR paclitaxel resistant cell line (Duan et al., 2005).

Purinergic receptor P2Y13 (P2RY13) is a G-protein-coupled receptor. Upon EMT induction due to EGF and hypoxia induction in breast cancer cells, P2RY13 was found downregulated (Davis et al., 2011; Azimi et al., 2016). In hepatocellular carcinoma, Maynard et al. reported better recurrence free survival for patients with increased expression of P2RY13 (Maynard et al., 2015).

Alpha-L-Fucosidase 1(FUCA1) is a lysosomal enzyme that plays an important role in the cleavage of terminal fructose residue. Its decreased expression has been related to several aggressive cancers including breast, colorectal and neuroblastoma cancers. In breast cancer, lymph node positive patients with decreased expression of FUCA1 showed decreased recurrence and cancer free survival (Bonin et al., 2018). Daves et al. reported an association between decreased expression of FUCA1 and increased metastasis in solid tumors (Daves et al., 2011). Ezawa et al. showed that p53 controls FUCA1 function and FUCA1 plays an important role EGFR signaling. Additionally, they reported FUCA1 as a survival biomarker for breast cancer (Ezawa et al., 2016). In TNBC patients, decreased

expression of FUCA1 was related to decreased survival as FUCA1 plays an important role in the modulation of glycosylation status (Cheng et al., 2015b).

Kinesin family member 13B (KIF13B) is a member of Kinesin 3 family. In breast cancer, E2 stimulation of breast cancer cell proliferation resulted in decreased KIF13B expression (Zou et al., 2014). The role of KIF13B is not fully explored in cancer. KIF13B is responsible for VEGFR2 trafficking and contributes in angiogenesis (Yamada et al., 2014). Park et al. reported KIF13B upregulation in MCF7 cells treated with paclitaxel for a longer period (Park et al., 2019).

Progesterone receptor (PGR, PR) is an important biomarker in breast cancer and is used clinically to subtype breast cancer. Its role in breast cancer plasticity is discussed in the previous section. Aurilio et al. reported that 46% of the tumors analyzed switched from PR positive to negative status upon metastasis (Aurilio et al., 2014). Similarly, Amir et al. showed that when primary and metastatic tissues were compared for PR, 40.4% of tissues showed a difference in PR status and 74% of these switched statuses were due to loss of PR positive status (Amir et al., 2012).

DIX domain containing 1 (DIXDC1) gene plays an important role in Wnt pathway. Additionally, its increased expression is related to good prognosis and metastasis repression in lung adenocarcinoma (Goodwin et al., 2014; Mao et al., 2017). Glioma cells showed decreased proliferation and migration upon DIXDC1 knockdown (Chen et al., 2017). Moreover decreased expression of DIXDC1 was associated with decreased survival in both hepatocellular carcinoma (Zhou et al., 2016) and intestinal type gastric carcinoma (Wang et al., 2015).

Joining chain of multimeric IgA and IgM (JCHAIN, IGJ) was reported to be downregulated in ductal carcinoma *in situ* and invasive ductal carcinoma when compared with normal stroma (Loi et al., 2009). Boersma et al. reported that both tumor stroma and tumor epithelium, in inflammatory breast cancer subtype, show decreased expression of JCHAIN when compared with non-inflammatory breast cancer (Boersma et al., 2008).

#### **6.4.2. Bad prognostic genes**

In our analysis, 10 genes were selected as bad prognosis genes because increased expression of these genes was related to increased metastasis. Serine dehydratase (SDS) is responsible for the conversion of L-serine to pyruvate and ammonia. Its role in breast cancer metastasis has not been fully explored. Mehrmohamadi et al. reported that SDS expression in tumor samples was higher in breast cancer when compared with normal tissues (Mehrmohamadi et al., 2014).

N-Alpha-Acetyltransferase 10, NatA Catalytic Subunit (NAA10, ARD1) is associated with post-translational modifications. And its overexpression has been associated with increased breast cancer proliferation (Yu et al., 2009b). When compared with normal cancer adjacent tissue, NAA10 expression was higher in cancer tissue (Yu et al., 2009a). Yu et al. showed a positive correlation between NAA10 expression and lymph node metastasis in breast cancer (Yu et al., 2009a). Surprisingly Zeng et al. reported NAA10 as an inhibitor of breast cancer metastasis which is opposite of our results and other studies reported here (Zeng et al., 2014).

Centromere protein E (CENPE) is identified as a hallmark of cancer along with chromosomal misalignment. It is involved in cancer cell proliferation as it plays an important role in capturing and positioning the chromosomes to mitotic spindle (Yardimci et al., 2008; El-Arabey et al., 2018). In breast cancer, CENPE inhibition was related to cell death and tumor regression (Kung et al., 2014; Ohashi et al., 2015).

Argonaute-2 (AGO2) is responsible for the post-transcriptional downregulation of gene expression through guided small-RNAs. It is part of the Argonaute protein family with 3 other members. AGO2 is solely responsible for catalytic activity among its family and thus causes gene silencing along RISC complex (Meister, 2013; Fischer et al., 2015). In the literature, several studies have associated AGO2 role in cancer development (Voller et al., 2013; Zhang et al., 2019). AGO2 expression is increased in tissue when compared with cell lines which shows the involvement of the microenvironment. AGO2 expression is higher in HER2 and Basal subtype of breast cancer and is a marker of relapse free survival (Casey et al., 2019). Additionally, in a subset of ER-positive breast cancer subtype AGO2

is expression is increased and this expression is related to decreased survival. Moreover, AGO2 overexpression also plays an important role to endocrine therapy (Conger et al., 2016). Adams et al reported that AGO2 overexpression in breast cancer cells resulted in increased proliferation and expression of pro-oncogenic microRNAs (Adams et al., 2009).

E2F Transcription Factor 8 (E2F8) is a member of E2F family and is responsible in several biological processes including cellular proliferation, differentiation, cell apoptosis, DNA repair and cell cycle (Logan et al., 2005; Lv et al., 2017). E2F8 is responsible for cell cycle progression as it binds to E2F1 and decreases its expression and thus allows the cell to transition from G1 to S phase (Kim et al., 2020). This process is exploited by cancer cells with higher E2F8 levels, in which E2F8 transcriptionally activates CCNE1 and CCNE2 leading to transition from G1 to S phase. This results in an increased proliferation of these cells (Ye et al., 2016). Lino et al. reported that E2F8 is responsible for increased proliferative potential due to post-transcriptional regulation through RNA binding proteins in breast cancer (Lino et al., 2020). Li et al. associated higher expression of E2F8 with poor breast cancer prognosis when analyzing TCGA breast cancer dataset (Li et al., 2018a).

BUB1 Mitotic Checkpoint Serine/Threonine Kinase B (BUB1B) expression shows a clear correlation with the grade and stage of breast cancer patients. Lower-grade breast cancer patients show decreased expression of BUB1B when compared with higher-grade tumors. Additionally, increased expression of this gene was related with decreased metastasis free survival (Cai et al., 2019a). Mansouri et al. showed that BUB1B expression was significantly higher among metastatic patients when compared with non-metastatic breast cancer patients (Mansouri et al., 2016).

Transmembrane Protein 208 (TMEM208) is a highly conserved gene and its protein is localized on endoplasmic reticulum. Increased expression of TMEM208 is related with ER stress and autophagy inhibition (Zhao et al., 2013). Additionally, oxygen levels in cells affect TMEM208 transcription. Upon hypoxic conditions, HIF-1 $\alpha$  binds at the promoter site of TMEM208 resulting in increased expression (Lei et al., 2020). So this gene can be used bad prognostic marker of breast cancer as an increase in hypoxia results in increased

invasion potential of cancer cells and this mechanism is controlled by HIF-1 $\alpha$  (Ren et al., 2015).

Proteasome 26S Subunit, Non-ATPase 2 (PSMD2) is responsible for protein part of proteasome subunit of the ubiquitin proteasome system (UPS). PSMD2 is highly overexpressed in breast cancer when compared with normal adjacent tissue by almost 3-fold (Deng et al., 2007). Increased expression of this gene is associated with shorter overall and distant metastasis free survival. Tomida et al. included this gene in their metastatic signature (Tomida et al., 2007). Additionally, Li et al. reported that knockdown of PSMD2 resulted in decreased proliferation and increased apoptosis (Li et al., 2018b).

Tubulin Alpha 1b (TUBA1B) belongs to the tubulin family which is responsible for major components of microtubules. TUBA1B is responsible for major biological processes including cell adhesion, migration, and division. Its role in breast cancer metastasis is not explored but Lu et al. reported increased expression of TUBA1B with decreased survival in hepatic cancer (Lu et al., 2013a).

Ubiquitin Conjugating Enzyme E2 A (UBE2A, RAD6A) is responsible in protein metabolism as a ubiquitin conjugating enzyme. Several studies have associated increased expression of UBE2A with aggressive breast cancer and metastasis (Shekhar et al., 2002) (Haynes et al., 2016). Its expression is also associated with cancer stem cell phenotype in ovarian cancer. Thus leading to disease recurrence and metastasis (Somasagara et al., 2017).

In conclusion, our identified metastatic biomarkers are good candidates for validation in future studies.

#### **6.4.3. Prognostic importance of hematopoietic cells for metastasis**

We used Cibersort tool to quantify twenty-two hematopoietic cells in breast tumors and checked if these values can be used to as metastatic prognostic markers for breast cancer metastasis. These cells did not show any statistical significance across eight datasets consistently. But their prognostic direction was found the same as previously described in the literature. Naive B cells showed good prognosis in 5/8 datasets. Zehran et al. reported

that with increase in grade of breast cancer patients, number of naïve B cells decrease (Zahran et al., 2020). Memory B cells showed good prognosis in 6 datasets. Garaud et al. reported that tumor infiltrating B cells play a critical role in generating humoral immune response and different B cell differentiation states, memory B cells show a consistent increase across breast cancer (Garaud et al., 2019). Plasma cells showed good prognosis in the majority of datasets and Yeong et al. related increases plasma cell levels with improved prognosis in TNBC (Yeong et al., 2018). CD8 T cells showed a better prognosis for breast cancer patients in 5 datasets and one of these was statistically significant. Mahmoud et al. reported that increased CD8 T cells is a favorable and independent prognostic marker for breast cancer (Mahmoud et al., 2011). CD4 naïve T cells also showed good prognosis but Su et al. reported that inhibition of CD4 T cells recruitment to breast cancer results in decreased tumor progression (Su et al., 2017). CD4 memory resting T cells and regulatory T cells showed bad prognosis. Reported that regulatory B cells play an important role in lung metastasis in metastasis as they are responsible for resting T cells to regulatory T cells and thus cause immune escape (Olkhanud et al., 2011). CD4 memory activated (Savas et al., 2018), follicular helper (Gu-Trantien et al., 2013) and gamma delta T cells showed good prognosis in 7 datasets. Surprisingly in literature tumor infiltration by gamma delta T cells has been associated with worse prognosis (Ma et al., 2012).

M1 macrophages showed good prognosis and M2 macrophages showed bad prognosis as expected in majority of datasets. But M0 showed an overall bad prognosis which is opposite to what was reported in the literature previously (Edechi et al., 2019). Resting mast cells showed good prognosis (7/8 datasets) and activated mast cells showed worse survival (6/8 datasets). The role of mast cells in breast cancer is not completely understood and further studies are needed (Aponte-López et al., 2018). Activated dendritic cells were associated with bad prognosis in 6/8 datasets.

Lastly, eosinophils showed good prognosis (Sakkal et al., 2016) while neutrophils showed bad prognosis in majority of the datasets (Coffelt et al., 2015).

In conclusion, although these cells did not show any significant prognostic association but their HR direction was found the same as previously described in the literature in majority

of datasets. Cibersort tool is still being refined and developed further and this analysis if repeated, after certain developments to the tool, will result in significant survival associations.

### **6.5. Prognostic importance of INHBA and MLANA in melanoma**

To examine the effects of EMT on cancer survival and metastasis in another cancer, role of INHBA and MLANA in melanoma was further evaluated. INHBA is a mesenchymal marker and MLANA is an epithelial marker for EMT. In our previous study, we have shown that these two genes collectively can classify melanoma in different subtypes such as stem/non-stem, invasive/proliferative and mesenchymal/epithelial (Senses et al., 2017). To identify if these genes can be used as prognostic biomarkers for melanoma, we performed the survival analysis in TCGA, GSE65904, GSE59455 and GSE54467 datasets for endpoints such as overall survival, DMFS and DSS. In all cases, cox p regression analyses performed yielded no significant results for MLANA and INHBA but using a cutoff selected through LRMC analysis to divide patients into High and Low expression groups, yielded significant relation between genes (MLANA and INHBA) and survival data. INHBA being a mesenchymal, invasive and stem cell marker was consistently found as good prognosis marker in melanoma patients and MLANA which is a epithelial, proliferative and non-stem marker was found consistently across datasets as bad prognostic gene. These analyses are in line with previous data regarding melanoma survival. Sinneberg et al. showed that INHBA was significantly overexpressed in the invasive phenotype when compared with the proliferative phenotype. Additionally, survival analysis showed INHBA as good prognosis marker similar to our analysis (Sinneberg et al., 2018). Luo et al reported PGC1 $\alpha$  as a metastasis suppressor gene. Additionally, they used INHBA as a member for their PGC1 $\alpha$  regulated metastasis/invasion associated gene signature (Luo et al., 2016). Gorlov et al. reported that INHBA plays a role in melanoma (Gorlov et al., 2018). Additionally, Jiang et al. showed that increased expression of INHBA is associated with increased overall and disease free survival in diffuse large B-cell lymphoma (Jiang et al., 2018). Surprisingly, INHBA overexpression was associated with poor survival in gastric (Katayama et al., 2017), lung adenocarcinoma (Seder et al.,

2009) and colorectal cancer (Okano et al., 2013) suggesting that INHBA's role as a prognostic marker is cancer specific.

MLANA has been reported as a biomarker to differentiate between invasive and proliferative melanoma subtypes (Cheng et al., 2015a). In line with our results, where we identified MLANA as a bad prognostic marker for melanoma, Aya-Bonilla et al. compared circulating tumor cells (CTCs) in melanoma patients before and after therapy. CTC scores were observed to be increased in these patients along with MLANA transcript number (Aya-Bonilla et al., 2020). MLANA plays an important role in melanoma metastasis as cancer cells overexpressing MLANA enter the blood circulation and become CTCs which then give rise to metastasis (Clawson et al., 2012). Role of MLANA-positive CTCs has been studied extensively (Curry et al., 1999; Kitago et al., 2009; Fusi et al., 2010). Koyanagi et al. reported that after therapy increased MLANA expression in blood specimens, was significantly associated with decreased RFS survival (Koyanagi et al., 2005). Raskin et al. identified increased expression of MLANA in primary cutaneous melanoma when compared with normal skin (Raskin et al., 2013). Increased antibody levels against MLANA was associated with better survival in melanoma patients (Fässler et al., 2019). Berset et al. showed that in addition to being a diagnostic marker, MLANA can be used to refine melanoma patients for prognostic importance in malignant melanoma (Berset et al., 2001). Gadiot et al. showed that 90% of their analyzed melanoma metastases were positive for MLANA by IHC (Gadiot et al., 2011). Reid et al. reported a multimarker qPCR based assay to detect CTCs in advanced melanoma patients. As expected MLANA showed high prognostic significance in terms of disease recurrence even when patients were considered disease-free (Reid et al., 2013).

In conclusion, this analysis showed that high expression of INHBA and MLANA is associated with increased and decreased survival of melanoma patients and these markers can be used as prognostic biomarkers for melanoma patients. Moreover, these findings in melanoma supported our results in breast cancer with CNCL gene list.

## 7. Future Perspectives

### 7.1. Classification of Breast cancer based upon stemness and EMT

Breast cancer is a heterogeneous disease that is responsible for the highest mortality in women. Breast cancer contains cancer stem cells which also show mesenchymal and drug-resistant properties. Here we identified CNCL gene list which can classify breast tumors based on stem cell-like and EMT like features into CS/M and NS/E phenotypes. We validated this in breast cancer cell lines, primary cancer cells and breast cancer tumor tissues *in silico*. Moreover, CNCL clusters drug-resistant cells separately from drug-sensitive cells for traditional drugs such as doxorubicin and paclitaxel. Additionally, when breast cancer cells were treated with drugs to generate resistant cells, granting them CS/M properties, CNCL could identify this dynamic shift quantitatively. We also identified drugs such as midostaurin which can selectively target CS/M phenotype and lapatinib can target NS/E phenotype of breast cancer. We validated these findings *in vitro* as well. CNCL gene expression qPCR data was highly correlated with CCLE microarray data and cell lines were clustered in the same CS/M and NS/E groups. Midostaurin and lapatinib sensitivity towards CS/M and NS/E phenotype were validated as well and when MDA-MB-157 cells were grown in 3D cell culture, they became more sensitive to midostaurin and resistant to lapatinib. Similarly, upon ZEB1 and SNAI2 knockdown in MDA-MB-157, cells became more epithelial leading to increased resistance to midostaurin and sensitivity to lapatinib. Although CNCL did not show any prognostic importance for patients treated with adjuvant therapy, it showed prognostic significance for patients treated with paclitaxel in the neoadjuvant setting. So CNCL can be used as a prognostic marker for these patients. Surprisingly most of the patient datasets treated with adjuvant therapy upon survival analysis, showed that patients with NS/E phenotype show worse survival when compared with CS/M phenotype patients. We believe this difference in prognosis for adjuvant and neoadjuvant therapy is because, in the neoadjuvant setting, exposure of cancer cells with drugs makes them resistant to therapy and mesenchymal, which results in more aggressive tumors. While in adjuvant setting the most of the tumor is removed in surgery and patients with increased proliferative potential tumors show worse survival. As discussed in the last section, recently in literature it has been shown that cancer cells with a combination of

epithelial and mesenchymal properties show an increased tendency to metastasize and generate aggressive tumors when compared with cells that show either epithelial or mesenchymal features. So, in the future, such gene lists should be identified which can classify breast cancer into different intermediary EMT stages as opposed to just binary classification.

## **7.2. Chemosensitivity markers of paclitaxel, doxorubicin and cisplatin in breast cancer**

In breast cancer, doxorubicin, paclitaxel and cisplatin are commonly used to treat patients who need chemotherapy. To identify biomarkers for these drugs' sensitivity, we correlated breast cancer cell lines gene expression data with their respective drug cytotoxicity data. For paclitaxel, we identified 20 potential biomarkers and these markers were used to generate two gene linear regression models. These two gene regression models are very promising as the gene expression of just two genes present in each model, can predict sensitivity to the drug for breast cancer cells when used in the linear regression formula. For paclitaxel 17 different models were generated and out of these only 3 models were selected as they were validated in 3 different datasets. Similarly, we identified 16 genes as potential biomarkers for cisplatin and 9 models were generated with regression analysis. We selected 2 cisplatin models as they were validated in 2 different datasets. For doxorubicin, 19 genes were selected as potential biomarkers and 10 linear regression models were generated. Out of these, only 2 models were selected as these could be validated in 4 out of 5 datasets. The two gene models for three drugs have the potential to predict cytotoxicity in cancer cell lines using gene expression data. But additional validation is required. For our future analysis, more datasets will be used to further validate our models and additionally, these models will also be validated *in vitro* by using cell lines that were not part of previous validations. Moreover, as these models can predict drug response, they also might show the prognostic significance for which survival analysis using multiple patient datasets will be performed *in silico*.

## **7.3. Breast cancer plasticity**

Breast cancer is classified based on the expression of IHC markers clinically and it is molecularly subtyped by PAM50. We evaluated the plasticity of breast cancer in response

to drug treatment or metastasis. We showed that PAM50 subtyping is highly plastic as even different biopsy collection methods such as core needle (stroma rich) and fine needle (stroma poor) biopsies can result in different molecular subtyping for certain patients. Additionally, we compared the molecular subtype of breast cancer patients before and after neoadjuvant chemotherapy and found that certain tumors can switch their subtype in response to treatment. Similar results were observed when primary and metastatic tissues' subtype were compared. Different subtypes of breast tumors show a different switch rate as LumA showed the highest plasticity while Basal subtype was most stable in response to drug treatment or metastasis. In breast cancer, as patients are treated according to their specific subtype such as LumA tumors are mostly treated with hormonal therapy and basal patients are treated with chemotherapy. So it is highly imperative that patients who show decreased response to therapies or show metastasis formation, should be re-evaluated for their subtype and treatment plan should be updated if such subtype switch is found. We could not perform any pathway analysis to explain the biology behind this switch. But we plan to perform such analysis in the future where patients' tumors will be compared before and after chemotherapy for gene expression to discover the biological pathways responsible for such switching.

#### **7.4. Biomarkers for metastasis and the role of hematopoietic cells in metastasis**

Breast cancer mortality rate has decreased over the last 20 years but its metastasis remains responsible for a high number of deaths as only 5% of patients with metastasis show long term survival. We identified 20 gene signature which can be used to determine the prognosis of such patients. Our signature includes 10 genes whose high expression is associated with good prognosis and 10 genes whose increased expression is associated with bad prognosis of breast cancer patients with metastasis. We used 8 different datasets to identify this signature and each gene was at least significant in at least 5/8 datasets. For future studies, we plan to validate this signature *ex vivo*. Additional analysis will also be performed to check if this signature can be used for the patients with lymph node metastasis. Additionally, we used Cibersort tool to quantify 22 hematopoietic cells' percentages in breast tumors using gene expression datasets. We further evaluated the prognostic importance of these cells as tumor microenvironment plays an important role in

breast cancer metastasis. Unfortunately, most of the cells did not show any prognostic significance. Importantly, most of the cells showed the same HR as previously reported in the literature in the majority of datasets. Cibersort tool is still being refined and developed further and this analysis if repeated, after certain developments to tool, will result in significant survival associations.

### **7.5. Role of INHBA and MLANA in melanoma**

In breast cancer, we observed that NS/E phenotype showed worse survival when compared with CS/M phenotype in the majority of datasets. We wanted to validate these findings in another cancer. Previously we have shown that INHBA (mesenchymal marker) and MLANA (epithelial marker) can stratify melanoma patients into invasive/proliferative, stem/non-stem and mesenchymal/epithelial groups. We used these genes to determine whether epithelial phenotype of melanoma shows better/worse survival when compared to mesenchymal phenotype. High expression of INHBA was associated with increased survival and high expression of MLANA was associated with decreased survival of melanoma patients. Thus, validating our results in breast cancer that epithelial phenotype shows worse survival when compared with mesenchymal phenotype. In the future, we plan to identify a pan-cancer gene signature that can be used to not only identify epithelial and mesenchymal phenotypes but also the stages in between EMT transition.

## 8. References

- Abu Samaan, T.M., Samec, M., Liskova, A., Kubatka, P., and Büsselberg, D. (2019). Paclitaxel's Mechanistic and Clinical Effects on Breast Cancer. *Biomolecules* 9(12), 789.
- AbuHammad, S., and Zihlif, M. (2013). Gene expression alterations in doxorubicin resistant MCF7 breast cancer cell line. *Genomics* 101(4), 213-220.
- Adams, A.L., Eltoum, I., Krontiras, H., Wang, W., and Chhieng, D.C. (2008). The effect of neoadjuvant chemotherapy on histologic grade, hormone receptor status, and HER2/neu status in breast carcinoma. *The breast journal* 14(2), 141-146.
- Adams, B.D., Claffey, K.P., and White, B.A. (2009). Argonaute-2 expression is regulated by epidermal growth factor receptor and mitogen-activated protein kinase signaling and correlates with a transformed phenotype in breast cancer cells. *Endocrinology* 150(1), 14-23. doi: 10.1210/en.2008-0984.
- Akbani, R., Akdemir, K.C., Aksoy, B.A., Albert, M., Ally, A., Amin, S.B., et al. (2015). Genomic classification of cutaneous melanoma. *Cell* 161(7), 1681-1696.
- Akbar, M.W. (2014). *Characterization of chemosensitivity profiles of breast cancer cell lines, with and without stem cell like features= Kök-hücre özelliği olan ve olmayan meme kanseri hücre hatlarının ilaç hassasiyet profillerinin tanımlanması*. Bilkent University.
- Akbar, M.W., Isbilen, M., Belder, N., Canli, S.D., Kucukkaraduman, B., Turk, C., et al. (2020a). A stemness and EMT based gene expression signature identifies phenotypic plasticity and is a predictive but not prognostic biomarker for breast cancer. *Journal of Cancer* 11(4), 949.
- Akbar, M.W., Isbilen, M., Kuckkaraduman, B., Canli, S.D., Dedeoglu, E., Azizolli, S., et al. (2020b). Plasticity as an Underlying Mechanism of Tumor Heterogeneity in Breast Cancer. 2020.2003.2011.987511. doi: 10.1101/2020.03.11.987511 %J bioRxiv.
- Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences* 100(7), 3983-3988.
- Alborzinia, H., Can, S., Holenya, P., Scholl, C., Lederer, E., Kitanovic, I., et al. (2011). Real-time monitoring of cisplatin-induced cell death. *PloS one* 6(5), e19714.
- Amir, E., Miller, N., Geddie, W., Freedman, O., Kassam, F., Simmons, C., et al. (2012). Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *Journal of clinical oncology* 30(6), 587.
- Anand, A.S., and Velayudhan, S.T. (2016). Discordance of Estrogen & Progesterone Receptors After Neoadjuvant Chemotherapy in Breast Cancer-an Indian Study. *Indian journal of surgical oncology* 7(3), 316-319.
- Aponte-López, A., Fuentes-Pananá, E.M., Cortes-Muñoz, D., and Muñoz-Cruz, S. (2018). Mast cell, the neglected member of the tumor microenvironment: role in breast cancer. *Journal of immunology research* 2018.
- Arozarena, I., and Wellbrock, C. (2019). Phenotype plasticity as enabler of melanoma progression and therapy resistance. *Nature Reviews Cancer* 19(7), 377-391.
- Ataseven, B., Gologan, D., Gunesch, A., Kehl, V., Hoegel, B., Beer, M., et al. (2012). HER2/neu, topoisomerase 2a, estrogen and progesterone receptors: discordance between primary breast cancer and metastatic axillary lymph node in expression and amplification characteristics. *Breast Care* 7(6), 465-470.
- Aurilio, G., Disalvatore, D., Pruneri, G., Bagnardi, V., Viale, G., Curigliano, G., et al. (2014). A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. *European journal of cancer* 50(2), 277-289.
- Austin, C.J., Kahlert, J., Kassiou, M., and Rendina, L.M. (2013). The translocator protein (TSPO): a novel target for cancer chemotherapy. *The international journal of biochemistry & cell biology* 45(7), 1212-1216.

- Aya-Bonilla, C.A., Morici, M., Hong, X., McEvoy, A.C., Sullivan, R.J., Freeman, J., et al. (2020). Detection and prognostic role of heterogeneous populations of melanoma circulating tumour cells. *British Journal of Cancer* 122(7), 1059-1067.
- Azimi, I., Beilby, H., Davis, F.M., Marcial, D.L., Kenny, P.A., Thompson, E.W., et al. (2016). Altered purinergic receptor-Ca<sup>2+</sup> signaling associated with hypoxia-induced epithelial-mesenchymal transition in breast cancer cells. *Molecular oncology* 10(1), 166-178.
- Bai, J., Zhang, J., Wu, J., Shen, L., Zeng, J., Ding, J., et al. (2010). JWA regulates melanoma metastasis by integrin  $\alpha$  V  $\beta$  3 signaling. *Oncogene* 29(8), 1227-1237.
- Bai, X., Ni, J., Beretov, J., Graham, P., and Li, Y. (2018). Cancer stem cell in breast cancer therapeutic resistance. *Cancer Treat Rev* 69, 152-163. doi: 10.1016/j.ctrv.2018.07.004.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391), 603-607. doi: 10.1038/nature11003.
- Bartucci, M., Dattilo, R., Moriconi, C., Pagliuca, A., Mottolise, M., Federici, G., et al. (2015). TAZ is required for metastatic activity and chemoresistance of breast cancer stem cells. *Oncogene* 34(6), 681-690.
- Battula, V.L., Shi, Y., Evans, K.W., Wang, R.-Y., Spaeth, E.L., Jacamo, R.O., et al. (2012). Ganglioside GD2 identifies breast cancer stem cells and promotes tumorigenesis. *The Journal of clinical investigation* 122(6), 2066-2078.
- Bell, R., Barraclough, R., and Vasieva, O. (2017). Gene expression meta-analysis of potential metastatic breast cancer markers. *Current molecular medicine* 17(3), 200-210.
- Bensimon, J., Biard, D., Paget, V., Goislar, M., Morel-Altmeier, S., Konge, J., et al. (2016). Forced extinction of CD24 stem-like breast cancer marker alone promotes radiation resistance through the control of oxidative stress. *Molecular carcinogenesis* 55(3), 245-254.
- Berset, M., Cerottini, J.P., Guggisberg, D., Romero, P., Burri, F., Rimoldi, D., et al. (2001). Expression of melan-a/MART-1 antigen as a prognostic factor in primary cutaneous melanoma. *International journal of cancer* 95(1), 73-77.
- Bhat-Nakshatri, P., Appaiah, H., Ballas, C., Pick-Franke, P., Goulet, R., Jr., Badve, S., et al. (2010). SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24-phenotype. *BMC Cancer* 10, 411. doi: 10.1186/1471-2407-10-411.
- Bhat-Nakshatri, P., Goswami, C.P., Badve, S., Sledge Jr, G.W., and Nakshatri, H. (2013). Identification of FDA-approved drugs targeting breast cancer stem cells along with biomarkers of sensitivity. *Scientific reports* 3, 2530.
- Bianchini, G., Qi, Y., Alvarez, R.H., Iwamoto, T., Coutant, C., Ibrahim, N.K., et al. (2010). Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and-negative cancers. *J Clin Oncol* 28(28), 4316-4323.
- Boersma, B.J., Reimers, M., Yi, M., Ludwig, J.A., Luke, B.T., Stephens, R.M., et al. (2008). A stromal gene signature associated with inflammatory breast cancer. *International journal of cancer* 122(6), 1324-1332.
- Bonin, S., Parascandolo, A., Aversa, C., Barbazza, R., Tsuchida, N., Castellone, M.D., et al. (2018). Reduced expression of  $\alpha$ -L-Fucosidase-1 (FUCA-1) predicts recurrence and shorter cancer specific survival in luminal B LN+ breast cancer patients. *Oncotarget* 9(20), 15228.
- Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., et al. (2009). Genes that mediate breast cancer metastasis to the brain. *Nature* 459(7249), 1005-1009. doi: 10.1038/nature08021.
- Broom, R.J., Tang, P.A., Simmons, C., Bordeleau, L., Mulligan, A.M., O'MALLEY, F.P., et al. (2009). Changes in estrogen receptor, progesterone receptor and Her-2/neu status with time: discordance rates between primary and metastatic breast cancer. *Anticancer research* 29(5), 1557-1562.

- Brown, G.A., McPherson, J.P., Gu, L., Hedley, D.W., Toso, R., Deuchars, K.L., et al. (1995). Relationship of DNA topoisomerase II $\alpha$  and  $\beta$  expression to cytotoxicity of antineoplastic agents in human acute lymphoblastic leukemia cell lines. *Cancer research* 55(1), 78-82.
- Budden, T., Davey, R.J., Vilain, R.E., Ashton, K.A., Braye, S.G., Beveridge, N.J., et al. (2016). Repair of UVB-induced DNA damage is reduced in melanoma due to low XPC and global genome repair. *Oncotarget* 7(38), 60940.
- Butti, R., Gunasekaran, V.P., Kumar, T.V., Banerjee, P., and Kundu, G.C. (2019). Breast cancer stem cells: Biology and therapeutic implications. *The international journal of biochemistry & cell biology* 107, 38-52.
- Byrski, T., Gronwald, J., Huzarski, T., Grzybowska, E., Budryk, M., Stawicka, M., et al. (2010). Pathologic complete response rates in young women with BRCA1-positive breast cancers after neoadjuvant chemotherapy. *Journal of Clinical Oncology* 28(3), 375-379.
- Cabioglu, N., Yazici, M.S., Arun, B., Broglio, K.R., Hortobagyi, G.N., Price, J.E., et al. (2005). CCR7 and CXCR4 as novel biomarkers predicting axillary lymph node metastasis in T1 breast cancer. *Clinical Cancer Research* 11(16), 5686-5693.
- Cai, J., Guan, H., Fang, L., Yang, Y., Zhu, X., Yuan, J., et al. (2013). MicroRNA-374a activates Wnt/ $\beta$ -catenin signaling to promote breast cancer metastasis. *The Journal of clinical investigation* 123(2).
- Cai, Y., Mei, J., Xiao, Z., Xu, B., Jiang, X., Zhang, Y., et al. (2019a). Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. *Hereditas* 156, 20. doi: 10.1186/s41065-019-0096-6.
- Cai, Y., Mei, J., Xiao, Z., Xu, B., Jiang, X., Zhang, Y., et al. (2019b). Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. *Hereditas* 156(1), 20.
- Calcagno, A.M., Salcido, C.D., Gillet, J.P., Wu, C.P., Fostel, J.M., Mumau, M.D., et al. (2010). Prolonged drug selection of breast cancer cells and enrichment of cancer stem cell characteristics. *J Natl Cancer Inst* 102(21), 1637-1652. doi: 10.1093/jnci/djq361.
- Carter, D., Dillon, D.C., Reynolds, L.D., Retter, M.W., Fanger, G., Molesh, D.A., et al. (2003). Serum antibodies to lipophilin B detected in late stage breast cancer patients. *Clinical cancer research* 9(2), 749-754.
- Casey, M.C., Prakash, A., Holian, E., McGuire, A., Kalinina, O., Shalaby, A., et al. (2019). Quantifying Argonaute 2 (Ago2) expression to stratify breast cancer. *BMC Cancer* 19(1), 712. doi: 10.1186/s12885-019-5884-x.
- Cava, C., Bertoli, G., Ripamonti, M., Mauri, G., Zoppis, I., Della Rosa, P.A., et al. (2014). Integration of mRNA expression profile, copy number alterations, and microRNA expression levels in breast cancer to improve grade definition. *PloS one* 9(5).
- Cejalvo, J.M., de Dueñas, E.M., Galván, P., García-Recio, S., Gasión, O.B., Paré, L., et al. (2017). Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer. *Cancer research* 77(9), 2213-2221.
- Chakrabarti, R., Hwang, J., Blanco, M.A., Wei, Y., Lukačičin, M., Romano, R.-A., et al. (2012). Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nature cell biology* 14(11), 1212-1222.
- Chakrabarti, R., Wei, Y., Hwang, J., Hang, X., Blanco, M.A., Choudhury, A., et al. (2014).  $\Delta$ Np63 promotes stem cell activity in mammary gland development and basal-like breast cancer by enhancing Fzd7 expression and Wnt signalling. *Nature cell biology* 16(10), 1004-1015.
- Chan, T.-S., Hsu, C.-C., Pai, V.C., Liao, W.-Y., Huang, S.-S., Tan, K.-T., et al. (2016). Metronomic chemotherapy prevents therapy-induced stromal activation and induction of tumor-initiating cells. *Journal of Experimental Medicine* 213(13), 2967-2988.
- Chang, J.C., Wooten, E.C., Tsimelzon, A., Hilsenbeck, S.G., Gutierrez, M.C., Elledge, R., et al. (2003). Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *The Lancet* 362(9381), 362-369.

- Charafe-Jauffret, E., Ginestier, C., Iovino, F., Wicinski, J., Cervera, N., Finetti, P., et al. (2009). Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer research* 69(4), 1302-1313.
- Chen, D., Bhat-Nakshatri, P., Goswami, C., Badve, S., and Nakshatri, H. (2013a). ANTXR1, a stem cell-enriched functional biomarker, connects collagen signaling to cancer stem-like cells and metastasis in breast cancer. *Cancer Res* 73(18), 5821-5833. doi: 10.1158/0008-5472.CAN-13-1080.
- Chen, J., Shen, C., Shi, J., Shen, J., Chen, W., Sun, J., et al. (2017). Knockdown of DIXDC1 inhibits the proliferation and migration of human glioma cells. *Cellular and molecular neurobiology* 37(6), 1009-1019.
- Chen, R., Li, A., Zhu, T., Li, C., Liu, Q., Chang, H.C., et al. (2005). JWA—A novel environmental-responsive gene, involved in estrogen receptor-associated signal pathway in MCF-7 and MDA-MB-231 breast carcinoma cells. *Journal of Toxicology and Environmental Health, Part A* 68(6), 445-456.
- Chen, R., Qiu, W., Liu, Z., Cao, X., Zhu, T., Li, A., et al. (2007). Identification of JWA as a novel functional gene responsive to environmental oxidative stress induced by benzo [a] pyrene and hydrogen peroxide. *Free Radical Biology and Medicine* 42(11), 1704-1714.
- Chen, W., Cai, F., Zhang, B., Barekati, Z., and Zhong, X.Y. (2013b). The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers. *Tumor Biology* 34(1), 455-462.
- Chen, W., Qin, Y., and Liu, S. (2018). Cytokines, breast cancer stem cells (BCSCs) and chemoresistance. *Clinical and translational medicine* 7(1), 27.
- Chen, X., Feng, J., Ge, Z., Chen, H., Ding, W., Zhu, W., et al. (2015). Effects of the JWA gene in the regulation of human breast cancer cells. *Molecular medicine reports* 11(5), 3848-3853.
- Chen, X., Lu, P., Wu, Y., Wang, D.-d., Zhou, S., Yang, S.-j., et al. (2016). MiRNAs-mediated cisplatin resistance in breast cancer. *Tumor Biology* 37(10), 12905-12913.
- Chen, Y., Xia, X., Wang, S., Wu, X., Zhang, J., Zhou, Y., et al. (2013c). High FAK combined with low JWA expression: clinical prognostic and predictive role for adjuvant fluorouracil–leucovorin–oxaliplatin treatment in resectable gastric cancer patients. *Journal of gastroenterology* 48(9), 1034-1044.
- Cheng, C.-C., Shi, L.-H., Wang, X.-J., Wang, S.-X., Wan, X.-Q., Liu, S.-R., et al. (2018). Stat3/Oct-4/c-Myc signal circuit for regulating stemness-mediated doxorubicin resistance of triple-negative breast cancer cells and inhibitory effects of WP1066. *International journal of oncology* 53(1), 339-348.
- Cheng, P.F., Shakhova, O., Widmer, D.S., Eichhoff, O.M., Zingg, D., Frommel, S.C., et al. (2015a). Methylation-dependent SOX9 expression mediates invasion in human melanoma cells and is a negative prognostic factor in advanced melanoma. *Genome biology* 16(1), 42.
- Cheng, T.-C., Tu, S.-H., Chen, L.-C., Chen, M.-Y., Chen, W.-Y., Lin, Y.-K., et al. (2015b). Down-regulation of  $\alpha$ -L-fucosidase 1 expression confers inferior survival for triple-negative breast cancer patients by modulating the glycosylation status of the tumor cell surface. *Oncotarget* 6(25), 21283.
- Chin, A.R., and Wang, S.E. (2014). Cytokines driving breast cancer stemness. *Molecular and cellular endocrinology* 382(1), 598-602.
- Cirenajwis, H., Ekedahl, H., Lauss, M., Harbst, K., Carneiro, A., Enoksson, J., et al. (2015). Molecular stratification of metastatic melanoma using gene expression profiling: Prediction of survival outcome and benefit from molecular targeted therapy. *Oncotarget* 6(14), 12297.
- Clara, J.A., Monge, C., Yang, Y., and Takebe, N. (2019). Targeting signalling pathways and the immune microenvironment of cancer stem cells—A clinical update. *Nature Reviews Clinical Oncology*, 1-29.
- Clawson, G.A., Kimchi, E., Patrick, S.D., Xin, P., Harouaka, R., Zheng, S., et al. (2012). Circulating tumor cells in melanoma patients. *PloS one* 7(7), e41052.

- Coffelt, S.B., Kersten, K., Doornebal, C.W., Weiden, J., Vrijland, K., Hau, C.-S., et al. (2015). IL-17-producing  $\gamma\delta$  T cells and neutrophils conspire to promote breast cancer metastasis. *Nature* 522(7556), 345-348.
- Conger, A.K., Martin, E.C., Yan, T.J., Rhodes, L.V., Hoang, V.T., La, J., et al. (2016). Argonaute 2 Expression Correlates with a Luminal B Breast Cancer Subtype and Induces Estrogen Receptor Alpha Isoform Variation. *Noncoding RNA* 2(3). doi: 10.3390/ncrna2030008.
- Cordenonsi, M., Zanonato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., et al. (2011). The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell* 147(4), 759-772.
- Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., et al. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A* 106(33), 13820-13825. doi: 10.1073/pnas.0905718106.
- Curry, B.J., Myers, K., and Hersey, P. (1999). MART-1 is expressed less frequently on circulating melanoma cells in patients who develop distant compared with locoregional metastases. *Journal of clinical oncology* 17(8), 2562-2562.
- Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403), 346-352. doi: 10.1038/nature10983.
- Da Costa, G.G., Gomig, T.H.B., Kaviski, R., Sousa, K.S., Kukolj, C., De Lima, R.S., et al. (2015). Comparative proteomics of tumor and paired normal breast tissue highlights potential biomarkers in breast cancer. *Cancer Genomics-Proteomics* 12(5), 251-261.
- Dasari, S., and Tchounwou, P.B. (2014). Cisplatin in cancer therapy: molecular mechanisms of action. *European journal of pharmacology* 740, 364-378.
- Daves, M.H., Hilsenbeck, S.G., Lau, C.C., and Man, T.-K. (2011). Meta-analysis of multiple microarray datasets reveals a common gene signature of metastasis in solid tumors. *BMC medical genomics* 4(1), 56.
- Davis, F.M., Kenny, P.A., Soo, E.T., van Denderen, B.J., Thompson, E.W., Cabot, P.J., et al. (2011). Remodeling of purinergic receptor-mediated Ca<sup>2+</sup> signaling as a consequence of EGF-induced epithelial-mesenchymal transition in breast cancer cells. *PloS one* 6(8).
- de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20(9), 1453-1454. doi: 10.1093/bioinformatics/bth078.
- Demirkol, S., Gomceli, I., Isbilen, M., Dayanc, B.E., Tez, M., Bostanci, E.B., et al. (2017). A Combined ULBP2 and SEMA5A Expression Signature as a Prognostic and Predictive Biomarker for Colon Cancer. *J Cancer* 8(7), 1113-1122. doi: 10.7150/jca.17872.
- Deng, S., Zhou, H., Xiong, R., Lu, Y., Yan, D., Xing, T., et al. (2007). Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast Cancer Res Treat* 104(1), 21-30. doi: 10.1007/s10549-006-9393-7.
- Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., et al. (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 13(11), 3207-3214. doi: 10.1158/1078-0432.ccr-06-2765.
- Dezső, Z., Oestreicher, J., Weaver, A., Santiago, S., Agoulnik, S., Chow, J., et al. (2014). Gene expression profiling reveals epithelial mesenchymal transition (EMT) genes can selectively differentiate eribulin sensitive breast cancer cells. *PLoS One* 9(8), e106131.
- Dieci, M.V., Barbieri, E., Piacentini, F., Ficarra, G., Bettelli, S., Dominici, M., et al. (2013). Discordance in receptor status between primary and recurrent breast cancer has a prognostic impact: a single-institution analysis. *Annals of oncology* 24(1), 101-108.

- DiMeo, T.A., Anderson, K., Phadke, P., Feng, C., Perou, C.M., Naber, S., et al. (2009). A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer research* 69(13), 5364-5373.
- Doebler, S.C., Sieuwerts, A.M., de Weerd, V., Stoop, H., Martens, J.W., and van Deurzen, C.H. (2017). Gene expression differences between ductal carcinoma in situ with and without progression to invasive breast cancer. *The American journal of pathology* 187(7), 1648-1655.
- Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nature reviews Molecular cell biology* 20(2), 69-84.
- Duan, Z., Lamendola, D.E., Duan, Y., Yusuf, R.Z., and Seiden, M.V. (2005). Description of paclitaxel resistance-associated genes in ovarian and breast cancer cell lines. *Cancer chemotherapy and pharmacology* 55(3), 277-285.
- Edechi, C.A., Ikeogu, N., Uzonna, J.E., and Myal, Y. (2019). Regulation of immunity in breast cancer. *Cancers* 11(8), 1080.
- Eger, A., Stockinger, A., Park, J., Langkopf, E., Mikula, M., Gotzmann, J., et al. (2004).  $\beta$ -Catenin and TGF  $\beta$  signalling cooperate to maintain a mesenchymal phenotype after FosER-induced epithelial to mesenchymal transition. 23(15), 2672-2680.
- Eichler, C., Flesch-Janys, D., Chang-Claude, J., Pantel, K., and Schwarzenbach, H. (2013). Deregulated serum concentrations of circulating cell-free microRNAs miR-17, miR-34a, miR-155, and miR-373 in human breast cancer development and progression. *Clinical chemistry* 59(10), 1489-1496.
- El-Arabey, A.A., Salama, S.A., and Abd-Allah, A.R. (2018). CENP-E as a target for cancer therapy: Where are we now? *Life Sci* 208, 192-200. doi: 10.1016/j.lfs.2018.07.037.
- Ellsworth, R., Valente, A., Field, L., Kane, J., Love, B., Tafral, L., et al. (2010). "Abstract P4-06-09: Genetic Signature Discriminating Metastatic from Non-Metastatic Small Tumors". AACR).
- Epis, M.R., Giles, K.M., Kalinowski, F.C., Barker, A., Cohen, R.J., and Leedman, P.J. (2012). Regulation of expression of deoxyhypusine hydroxylase (DOHH), the enzyme that catalyzes the activation of eIF5A, by miR-331-3p and miR-642-5p in prostate cancer cells. *Journal of Biological Chemistry* 287(42), 35251-35259.
- Erdem, G.U., Altundag, K., Ozdemir, N.Y., Sahin, S., Demirci, N.S., Karatas, F., et al. (2017). Comparative study of receptor discordance between primary and corresponding metastatic lesions in breast cancer. *J BUON* 22(2), 365-376.
- Ezawa, I., Sawai, Y., Kawase, T., Okabe, A., Tsutsumi, S., Ichikawa, H., et al. (2016). Novel p53 target gene FUCA 1 encodes a fucosidase and regulates growth and survival of cancer cells. *Cancer science* 107(6), 734-745.
- Fackler, M.J., Bujanda, Z.L., Umbricht, C., Teo, W.W., Cho, S., Zhang, Z., et al. (2014). Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. *Cancer research* 74(8), 2160-2170.
- Farnie, G., and Clarke, R.B. (2007). Mammary stem cells and breast cancer—role of Notch signalling. *Stem cell reviews* 3(2), 169-175.
- Fässler, M., Diem, S., Mangana, J., Ali, O.H., Berner, F., Bomze, D., et al. (2019). Antibodies as biomarker candidates for response and survival to checkpoint inhibitors in melanoma patients. *Journal for immunotherapy of cancer* 7(1), 50.
- Fillmore, C.M., and Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast cancer research* 10(2), 1-13.
- Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T., et al. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* 527(7579), 472-476.

- Floor, S., Van Staveren, W., Larsimont, D., Dumont, J.E., and Maenhaut, C. (2011). Cancer cells in epithelial-to-mesenchymal transition and tumor-propagating–cancer stem cells: distinct, overlapping or same populations. *Oncogene* 30(46), 4609-4621.
- Frisch, S.M., Schaller, M., and Cieply, B. (2013). Mechanisms that link the oncogenic epithelial–mesenchymal transition to suppression of anoikis. *Journal of cell science* 126(1), 21-29.
- Fujita, T., Kitaura, F., and Fujii, H. (2015). A critical role of the Thy28-MYH9 axis in B cell-specific expression of the Pax5 gene in chicken B cells. *PLoS One* 10(1).
- Fusi, A., Ochsenreither, S., Busse, A., Rietz, A., and Keilholz, U. (2010). Expression of the stem cell marker nestin in peripheral blood of patients with melanoma. *British Journal of Dermatology* 163(1), 107-114.
- Gadiot, J., Hooijkaas, A.I., Kaiser, A.D., van Tinteren, H., van Boven, H., and Blank, C. (2011). Overall survival and PD-L1 expression in metastasized malignant melanoma. *Cancer* 117(10), 2192-2201.
- Garaud, S., Buisseret, L., Solinas, C., Gu-Trantien, C., de Wind, A., Van den Eynden, G., et al. (2019). Tumor-infiltrating B cells signal functional humoral immune responses in breast cancer. *JCI insight* 4(18).
- Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483(7391), 570-575. doi: 10.1038/nature11005.
- Gjerdrum, C., Tiron, C., Høiby, T., Stefansson, I., Haugen, H., Sandal, T., et al. (2010). Axl is an essential epithelial-to-mesenchymal transition-induced regulator of breast cancer metastasis and patient survival. *Proceedings of the National Academy of Sciences* 107(3), 1124-1129.
- Gobbini, E., Ezzalfani, M., Dieras, V., Bachelot, T., Brain, E., Debled, M., et al. (2018). Time trends of overall survival among metastatic breast cancer patients in the real-life ESME cohort. *European Journal of Cancer* 96, 17-24.
- Goggins, W., Gao, W., and Tsao, H. (2004). Association between female breast cancer and cutaneous melanoma. *International journal of cancer* 111(5), 792-794.
- Gonzalez-Angulo, A.M., Iwamoto, T., Liu, S., Chen, H., Do, K.-A., Hortobagyi, G.N., et al. (2012). Gene expression, molecular class changes, and pathway analysis after neoadjuvant systemic therapy for breast cancer. *Clinical Cancer Research* 18(4), 1109-1119.
- Goodwin, J.M., Svensson, R.U., Lou, H.J., Winslow, M.M., Turk, B.E., and Shaw, R.J. (2014). An AMPK-independent signaling pathway downstream of the LKB1 tumor suppressor controls Snail1 and metastatic potential. *Molecular cell* 55(3), 436-450.
- Gorlov, I., Orlov, I., Ringelberg, C., Hernando, E., Ernstoff, M.S., Cheng, C., et al. (2018). Identification of gene expression levels in primary melanoma associated with clinically meaningful characteristics. *Melanoma research* 28(5), 380.
- Gradishar, W.J. (2012). Taxanes for the treatment of metastatic breast cancer. *Breast cancer: basic and clinical research* 6, BCBCR. S8205.
- Grimshaw, M.J., Cooper, L., Papazisis, K., Coleman, J.A., Bohnenkamp, H.R., Chiapero-Stanke, L., et al. (2008). Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. *Breast Cancer Research* 10(3), R52.
- Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., De Wind, A., et al. (2013). CD4+ follicular helper T cell infiltration predicts breast cancer survival. *The Journal of clinical investigation* 123(7), 2873-2892.
- Guo, Q., Cheng, K., Wang, X., Li, X., Yu, Y., Hua, Y., et al. (2020). Expression of HDAC1 and RBBP4 correlate with clinicopathologic characteristics and prognosis in breast cancer. *International Journal of Clinical and Experimental Pathology* 13(3), 563.
- Gupta, P.B., Onder, T.T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R.A., et al. (2009a). Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 138(4), 645-659. doi: 10.1016/j.cell.2009.06.034.

- Gupta, P.B., Onder, T.T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R.A., et al. (2009b). Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 138(4), 645-659.
- Gupta, P.B., Pastushenko, I., Skibinski, A., Blanpain, C., and Kuperwasser, C. (2019). Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* 24(1), 65-78. doi: 10.1016/j.stem.2018.11.011.
- Györfy, B., and Schäfer, R. (2009). Meta-analysis of gene expression profiles related to relapse-free survival in 1,079 breast cancer patients. *Breast cancer research and treatment* 118(3), 433-441.
- Han, W., Zhang, C., Cao, F.-y., Cao, F., Jiang, L., and Ding, H.-z. (2017). Prognostic and clinicopathological value of NM23 expression in patients with breast cancer: A systematic review and meta-analysis. *Current Problems in Cancer* 41(1), 80-93.
- Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., et al. (2019). Breast cancer. *Nature Reviews Disease Primers* 5(1), 66. doi: 10.1038/s41572-019-0111-2.
- Hatzis, C., Pusztai, L., Valero, V., Booser, D.J., Esserman, L., Lluch, A., et al. (2011). A genomic predictor of response and survival following taxane- anthracycline chemotherapy for invasive breast cancer. *Jama* 305(18), 1873-1881. doi: 10.1001/jama.2011.593.
- Haynes, B., Zhang, Y., Liu, F., Li, J., Petit, S., Kothayer, H., et al. (2016). Gold nanoparticle conjugated Rad6 inhibitor induces cell death in triple negative breast cancer cells by inducing mitochondrial dysfunction and PARP-1 hyperactivation: Synthesis and characterization. *Nanomedicine* 12(3), 745-757. doi: 10.1016/j.nano.2015.10.010.
- Heng, Y.J., Lester, S.C., Tse, G.M., Factor, R.E., Allison, K.H., Collins, L.C., et al. (2017). The molecular basis of breast cancer pathological phenotypes. *The Journal of pathology* 241(3), 375-391.
- Hii, L.-W., Chung, F.F.-L., Soo, J.S.-S., Tan, B.S., Mai, C.-W., and Leong, C.-O. (2020). Histone deacetylase (HDAC) inhibitors and doxorubicin combinations target both breast cancer stem cells and non-stem breast cancer cells simultaneously. *Breast cancer research and treatment* 179(3), 615-629.
- Hirata, T., Shimizu, C., Yonemori, K., Hirakawa, A., Kouno, T., Tamura, K., et al. (2009). Change in the hormone receptor status following administration of neoadjuvant chemotherapy and its impact on the long-term outcome in patients with primary breast cancer. *British journal of cancer* 101(9), 1529-1536.
- Ho, W., Comber, H., Hill, A., and Murphy, G. (2011). Malignant melanoma and breast carcinoma: a bidirectional correlation. *Irish journal of medical science* 180(4), 901-903.
- Hodgkinson, V.C., Eagle, G.L., Drew, P.J., Lind, M.J., and Cawthell, L. (2010). Biomarkers of chemotherapy resistance in breast cancer identified by proteomics: current status. *Cancer letters* 294(1), 13-24.
- Hoek, K.S., Eichhoff, O.M., Schlegel, N.C., Döbbeling, U., Kobert, N., Schaerer, L., et al. (2008). In vivo switching of human melanoma cells between proliferative and invasive states. *Cancer research* 68(3), 650-656.
- Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., et al. (2014). Drug resistance in cancer: an overview. *Cancers (Basel)* 6(3), 1769-1792. doi: 10.3390/cancers6031769.
- Hsu, Y.C., Chen, H.Y., Yuan, S., Yu, S.L., Lin, C.H., Wu, G., et al. (2013). Genome-wide analysis of three-way interplay among gene expression, cancer cell invasion and anti-cancer compound sensitivity. *BMC Med* 11, 106. doi: 10.1186/1741-7015-11-106.
- Iino, K., Mitobe, Y., Ikeda, K., Takayama, K.I., Suzuki, T., Kawabata, H., et al. (2020). RNA-binding protein NONO promotes breast cancer proliferation by post-transcriptional regulation of SKP2 and E2F8. *Cancer Sci* 111(1), 148-159. doi: 10.1111/cas.14240.

- Imani, S., Hosseinifard, H., Cheng, J., Wei, C., and Fu, J. (2016). Prognostic value of EMT-inducing transcription factors (EMT-TFs) in metastatic breast cancer: a systematic review and meta-analysis. *Scientific reports* 6(1), 1-10.
- Inao, T., Iida, Y., Moritani, T., Okimoto, T., Tanino, R., Kotani, H., et al. (2018). Bcl-2 inhibition sensitizes triple-negative human breast cancer cells to doxorubicin. *Oncotarget* 9(39), 25545.
- Isbilen, M., Mert Senses, K., and Osmay Gure, A.J.C.S.T.T. (2013). Predicting Chemotherapy Sensitivity Profiles for Breast Cancer Cell Lines with and Without Stem Cell-Like Features. *8*(3), 268-273.
- Islam, F., Gopalan, V., Smith, R.A., and Lam, A.K.-Y. (2015). Translational potential of cancer stem cells: A review of the detection of cancer stem cells and their roles in cancer recurrence and cancer treatment. *Experimental cell research* 335(1), 135-147.
- Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., et al. (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 66(21), 10292-10301. doi: 10.1158/0008-5472.can-05-4414.
- Iwamoto, T., Niikura, N., Ogiya, R., Yasojima, H., Watanabe, K.-i., Kanbayashi, C., et al. (2019). Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Scientific reports* 9(1), 1-8.
- Jayawardana, K., Schramm, S.J., Haydu, L., Thompson, J.F., Scolyer, R.A., Mann, G.J., et al. (2015). Determination of prognosis in metastatic melanoma through integration of clinico-pathologic, mutation, mRNA, microRNA, and protein information. *International journal of cancer* 136(4), 863-874.
- Jezequel, P., Loussouarn, D., Guerin-Charbonnel, C., Campion, L., Vanier, A., Gouraud, W., et al. (2015). Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res* 17, 43. doi: 10.1186/s13058-015-0550-y.
- Jiang, L., Si, T., Yu, M., Zeng, X., Morse III, H.C., Lu, Y., et al. (2018). The tumor suppressive role of inhibin  $\beta$  A in diffuse large B-cell lymphoma. *Leukemia & lymphoma* 59(5), 1202-1212.
- Jiang, X., Toyota, H., Yoshimoto, T., Takada, E., Asakura, H., and Mizuguchi, J. (2003). Anti-IgM-induced down-regulation of nuclear Thy28 protein expression in Ramos B lymphoma cells. *Apoptosis* 8(5), 509-519.
- Jong, N.N., and McKeage, M.J. (2014). Emerging roles of metal solute carriers in cancer mechanisms and treatment. *Biopharmaceutics & drug disposition* 35(8), 450-462.
- Kakarla, S., Song, X.T., and Gottschalk, S. (2012). Cancer-associated fibroblasts as targets for immunotherapy. *Immunotherapy* 4(11), 1129-1138. doi: 10.2217/imt.12.112.
- Kang, R., Xie, Y., Zeh, H.J., Klionsky, D.J., and Tang, D. (2019). Mitochondrial quality control mediated by PINK1 and PRKN: links to iron metabolism and tumor immunity. *Autophagy* 15(1), 172-173.
- Kao, K.J., Chang, K.M., Hsu, H.C., and Huang, A.T. (2011). Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* 11, 143. doi: 10.1186/1471-2407-11-143.
- Kasami, M., Uematsu, T., Honda, M., Yabuzaki, T., Sanuki, J., Uchida, Y., et al. (2008). Comparison of estrogen receptor, progesterone receptor and Her-2 status in breast cancer pre- and post-neoadjuvant chemotherapy. *The Breast* 17(5), 523-527.
- Kasper, M., Jaks, V., Fiaschi, M., and Toftgård, R. (2009). Hedgehog signalling in breast cancer. *Carcinogenesis* 30(6), 903-911.
- Katayama, Y., Oshima, T., Sakamaki, K., Aoyama, T., Sato, T., Masudo, K., et al. (2017). Clinical significance of INHBA gene expression in patients with gastric cancer who receive curative resection followed by adjuvant S-1 chemotherapy. *In vivo* 31(4), 565-571.
- Kim, B., Stephen, S.L., Hanby, A.M., Horgan, K., Perry, S.L., Richardson, J., et al. (2015). Chemotherapy induces Notch1-dependent MRP1 up-regulation, inhibition of which sensitizes breast cancer cells to chemotherapy. *BMC cancer* 15(1), 634.

- Kim, H.M., Lee, J., and Koo, J.S. (2017). Clinicopathological and prognostic significance of programmed death ligand-1 expression in breast cancer: a meta-analysis. *BMC cancer* 17(1), 690.
- Kim, K., Lu, Z., and Hay, E.D.J.C.b.i. (2002). Direct evidence for a role of  $\beta$ -catenin/LEF-1 signaling pathway in induction of EMT. 26(5), 463-476.
- Kim, L.K., Park, S.A., Eoh, K.J., Heo, T.H., Kim, Y.T., and Kim, H.J. (2020). E2F8 regulates the proliferation and invasion through epithelial-mesenchymal transition in cervical cancer. *Int J Biol Sci* 16(2), 320-329. doi: 10.7150/ijbs.37686.
- Kinoe, H., Yamanouchi, K., Kuba, S., Morita, M., Sakimura, C., Kanetaka, K., et al. (2018). Discordance of hormone receptor, human epidermal growth factor receptor-2, and Ki-67 between primary breast cancer and synchronous axillary lymph node metastasis. *JBUON* 23, S60-S6666.
- Kitago, M., Koyanagi, K., Nakamura, T., Goto, Y., Faries, M., O'day, S.J., et al. (2009). mRNA expression and BRAF mutation in circulating melanoma cells isolated from peripheral blood with high molecular weight melanoma-associated antigen-specific monoclonal antibody beads. *Clinical chemistry* 55(4), 757-764.
- Klein, M.E., Dabbs, D.J., Shuai, Y., Brufsky, A.M., Jankowitz, R., Puhalla, S.L., et al. (2013). Prediction of the Onco type DX recurrence score: use of pathology-generated equations derived by linear regression analysis. *Modern Pathology* 26(5), 658-664.
- Klinton, M., Buus, R., Cheang, M.C., Sheri, A., Smith, I.E., and Dowsett, M. (2016). Changes in Expression of Genes Representing Key Biologic Processes after Neoadjuvant Chemotherapy in Breast Cancer, and Prognostic Implications in Residual Disease. *J Exp Med* 22(10), 2405-2416. doi: 10.1084/jem.20151665
- 10.1158/1078-0432.ccr-15-1488.
- Kong, J., Leteurtre, F., Goisard, M., Biard, D., Morel-Altmeier, S., Vaurijoux, A., et al. (2018). Breast cancer stem cell-like cells generated during TGF $\beta$ -induced EMT are radioresistant. *Oncotarget* 9(34), 23519.
- Koyanagi, K., O'Day, S.J., Gonzalez, R., Lewis, K., Robinson, W.A., Amatruda, T.T., et al. (2005). Serial monitoring of circulating melanoma cells during neoadjuvant biochemotherapy for stage III melanoma: outcome prediction in a multicenter trial. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 23(31), 8057.
- Krebs, A.M., Mitschke, J., Laserra Losada, M., Schmalhofer, O., Boerries, M., Busch, H., et al. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat Cell Biol* 19(5), 518-529. doi: 10.1038/ncb3513.
- Kucukkaraduman, B., Turk, C., Fallacara, A.L., Isbilen, M., Senses, K.M., Ayyildiz, Z.O., et al. (2020). Predictive Gene Signature for Pyrazolopyrimidine Derivative c-Src Inhibitor 10a Sensitivity in Melanoma Cells. *ACS Medicinal Chemistry Letters* 11(5), 928-932.
- Kung, P.P., Martinez, R., Zhu, Z., Zager, M., Blasina, A., Rymer, I., et al. (2014). Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer. *Mol Cancer Ther* 13(8), 2104-2115. doi: 10.1158/1535-7163.Mct-14-0083-t.
- Kuzmanov, A., Hopfer, U., Marti, P., Meyer-Schaller, N., Yilmaz, M., and Christofori, G. (2014). LIM-homeobox gene 2 promotes tumor growth and metastasis by inducing autocrine and paracrine PDGF-B signaling. *Molecular oncology* 8(2), 401-416.
- Lagade, C., Vlasi, E., Della Donna, L., Dekmezian, C., and Pajonk, F. (2012). Radiation-induced reprogramming of breast cancer cells. *Stem cells* 30(5), 833-844.
- Lambert, A.W., Pattabiraman, D.R., and Weinberg, R.A. (2017). Emerging biological principles of metastasis. *Cell* 168(4), 670-691.
- Lei, P., Xiao, Y., Li, P., Xie, P., Wang, H., Huang, S., et al. (2020). hSnd2/TMEM208 is an HIF-1 $\alpha$ -targeted gene and contains a WH2 motif. *Acta Biochim Biophys Sin (Shanghai)*. doi: 10.1093/abbs/gmz157.

- Leong, C.-O., Vidnovic, N., DeYoung, M.P., Sgroi, D., and Ellisen, L.W. (2007). The p63/p73 network mediates chemosensitivity to cisplatin in a biologically defined subset of primary breast cancers. *The Journal of clinical investigation* 117(5), 1370-1380.
- Li, H., Zhu, Y., Burnside, E.S., Drukker, K., Hoadley, K.A., Fan, C., et al. (2016). MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *Radiology* 281(2), 382-391.
- Li, Q., and Shu, Y. (2014). Role of solute carriers in response to anticancer drugs. *Molecular and cellular therapies* 2(1), 1-14.
- Li, Y.-D., Lv, Z., Xie, H.-Y., and Zheng, S.-S. (2019). Retinoblastoma binding protein 4 up-regulation is correlated with hepatic metastasis and poor prognosis in colon cancer patients. *Hepatobiliary & Pancreatic Diseases International* 18(5), 446-451.
- Li, Y., Bai, W., and Zhang, L. (2020). The overexpression of CD80 and ISG15 are associated with the progression and metastasis of breast cancer by a meta-analysis integrating three microarray datasets. *Pathology & Oncology Research* 26(1), 443-452.
- Li, Y., Huang, J., Yang, D., Xiang, S., Sun, J., Li, H., et al. (2018a). Expression patterns of E2F transcription factors and their potential prognostic roles in breast cancer. *Oncol Lett* 15(6), 9216-9230. doi: 10.3892/ol.2018.8514.
- Li, Y., Huang, J., Zeng, B., Yang, D., Sun, J., Yin, X., et al. (2018b). PSMD2 regulates breast cancer cell proliferation and cell cycle progression by modulating p21 and p27 proteasomal degradation. *Cancer Lett* 430, 109-122. doi: 10.1016/j.canlet.2018.05.018.
- Li, Y., Zou, L., Li, Q., Haibe-Kains, B., Tian, R., Li, Y., et al. (2010). Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat Med* 16(2), 214-218. doi: 10.1038/nm.2090.
- Li, Z.-L., Chen, C., Yang, Y., Wang, C., Yang, T., Yang, X., et al. (2015). Gamma secretase inhibitor enhances sensitivity to doxorubicin in MDA-MB-231 cells. *International journal of clinical and experimental pathology* 8(5), 4378.
- Liang, Y., Song, X., Li, Y., Su, P., Han, D., Ma, T., et al. (2019). circKDM4C suppresses tumor progression and attenuates doxorubicin resistance by regulating miR-548p/PBLD axis in breast cancer. *Oncogene* 38(42), 6850-6866.
- Lindstrom, L.S., Karlsson, E., Wilking, U.M., Johansson, U., Hartman, J., Lidbrink, E.K., et al. (2012). Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol* 30(21), 2601-2608.
- Liu, J., Shen, J.-X., Wu, H.-T., Li, X.-L., Wen, X.-F., Du, C.-W., et al. (2018a). Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. *Discovery medicine* 25(139), 211-223.
- Liu, L., Greger, J., Shi, H., Liu, Y., Greshock, J., Annan, R., et al. (2009). Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. *Cancer research* 69(17), 6871-6878.
- Liu, L., Yang, L., Yan, W., Zhai, J., Pizzo, D.P., Chu, P., et al. (2018b). Chemotherapy Induces Breast Cancer Stemness in Association with Dysregulated Monocytosis. *Clin Cancer Res* 24(10), 2370-2382. doi: 10.1158/1078-0432.ccr-17-2545.
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., et al. (2014a). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem cell reports* 2(1), 78-91.
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., et al. (2014b). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* 2(1), 78-91. doi: 10.1016/j.stemcr.2013.11.009.
- Liu, S., and Wicha, M.S. (2010). Targeting breast cancer stem cells. *Journal of clinical oncology* 28(25), 4006.

- Liu, Z., Teng, L., Gao, L., Wang, H., Su, Y., and Li, J. (2018c). The role of eukaryotic translation initiation factor 5A-1 (eIF5A-1) gene in HPV 16 E6 induces cell growth in human cervical squamous carcinoma cells. *Biochemical and biophysical research communications* 504(1), 6-12.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4), 402-408. doi: 10.1006/meth.2001.1262.
- Logan, N., Graham, A., Zhao, X., Fisher, R., Maiti, B., Leone, G., et al. (2005). E2F-8: an E2F family member with a similar organization of DNA-binding domains to E2F-7. *Oncogene* 24(31), 5000-5004. doi: 10.1038/sj.onc.1208703.
- Loi, S., Haibe-Kains, B., Desmedt, C., Lallemand, F., Tutt, A.M., Gillet, C., et al. (2007). Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 25(10), 1239-1246. doi: 10.1200/jco.2006.07.1522.
- Loi, S., Pommey, S., Haibe-Kains, B., Beavis, P.A., Darcy, P.K., Smyth, M.J., et al. (2013). CD73 promotes anthracycline resistance and poor prognosis in triple negative breast cancer. *Proceedings of the National Academy of Sciences* 110(27), 11091-11096.
- Loi, S., Sotiriou, C., Haibe-Kains, B., Lallemand, F., Conus, N.M., Piccart, M.J., et al. (2009). Gene expression profiling identifies activated growth factor signaling in poor prognosis (Luminal-B) estrogen receptor positive breast cancer. *BMC medical genomics* 2(1), 37.
- Long, J., Ji, Z., Jiang, K., Wang, Z., and Meng, G. (2015). miR-193b modulates resistance to doxorubicin in human breast cancer cells by downregulating MCL-1. *BioMed research international* 2015.
- Lorusso, G., and Rüegg, C. (Year). "New insights into the mechanisms of organ-specific breast cancer metastasis", in: *Seminars in cancer biology*: Elsevier), 226-233.
- Lu, C., Zhang, J., He, S., Wan, C., Shan, A., Wang, Y., et al. (2013a). Increased alpha-tubulin1b expression indicates poor prognosis and resistance to chemotherapy in hepatocellular carcinoma. *Dig Dis Sci* 58(9), 2713-2720. doi: 10.1007/s10620-013-2692-z.
- Lu, J., Tang, Y., Cheng, Y., Zhang, G., Yip, A., Martinka, M., et al. (2013b). ING4 regulates JWA in angiogenesis and their prognostic value in melanoma patients. *British journal of cancer* 109(11), 2842-2852.
- Luo, C., Lim, J.-H., Lee, Y., Granter, S.R., Thomas, A., Vazquez, F., et al. (2016). A PGC1 $\alpha$ -mediated transcriptional axis suppresses melanoma metastasis. *Nature* 537(7620), 422-426.
- Luo, W., Schork, N.J., Marschke, K.B., Ng, S.-C., Hermann, T.W., Zhang, J., et al. (2011). Identification of polymorphisms associated with hypertriglyceridemia and prolonged survival induced by Bexarotene in treating non-small cell lung cancer. *Anticancer research* 31(6), 2303-2311.
- Lv, Y., Xiao, J., Liu, J., and Xing, F. (2017). E2F8 is a Potential Therapeutic Target for Hepatocellular Carcinoma. *J Cancer* 8(7), 1205-1213. doi: 10.7150/jca.18255.
- Ma, C., Zhang, Q., Ye, J., Wang, F., Zhang, Y., Wevers, E., et al. (2012). Tumor-infiltrating  $\gamma\delta$  T lymphocytes predict clinical outcome in human breast cancer. *The Journal of Immunology* 189(10), 5029-5036.
- Mahmoud, S.M., Paish, E.C., Powe, D.G., Macmillan, R.D., Grainge, M.J., Lee, A.H., et al. (2011). Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *Journal of clinical oncology* 29(15), 1949-1955.
- Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133(4), 704-715.
- Mansouri, N., Movafagh, A., Sayad, A., Heidary Pour, A., Taheri, M., Soleimani, S., et al. (2016). Targeting of BUB1b Gene Expression in Sentinel Lymph Node Biopsies of Invasive

- Breast Cancer in Iranian Female Patients. *Asian Pac J Cancer Prev* 17(S3), 317-321. doi: 10.7314/apjcp.2016.17.s3.317.
- Mao, X.Y., Lee, M.J., Zhu, J., Zhu, C., Law, S.M., and Snijders, A.M. (2017). Genome-wide screen identifies a novel prognostic signature for breast cancer survival. *Oncotarget* 8(8), 14003.
- Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., et al. (2016). Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* 164(1-2), 293-309. doi: 10.1016/j.cell.2015.11.062.
- Marinello, P.C., Panis, C., Silva, T.N.X., Binato, R., Abdelhay, E., Rodrigues, J.A., et al. (2019). Metformin prevention of doxorubicin resistance in MCF-7 and MDA-MB-231 involves oxidative stress generation and modulation of cell adaptation genes. *Scientific reports* 9(1), 1-11.
- Matrone, M.A., Whipple, R.A., Thompson, K., Cho, E.H., Vitolo, M.I., Balzer, E.M., et al. (2010). Metastatic breast tumors express increased tau, which promotes microtentacle formation and the reattachment of detached breast tumor cells. *Oncogene* 29(22), 3217-3227.
- Maynard, J.P., Lee, J.-S., Sohn, B.H., Yu, X., Lopez-Terrada, D., Finegold, M.J., et al. (2015). P2X3 purinergic receptor overexpression is associated with poor recurrence-free survival in hepatocellular carcinoma patients. *Oncotarget* 6(38), 41162.
- Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M.S., Ghosh, D., et al. (2005). Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer research* 65(24), 11259-11264.
- Mehrmohamadi, M., Liu, X., Shestov, A.A., and Locasale, J.W. (2014). Characterization of the usage of the serine metabolic network in human cancer. *Cell Rep* 9(4), 1507-1519. doi: 10.1016/j.celrep.2014.10.026.
- Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* 14(7), 447-459. doi: 10.1038/nrg3462.
- Mellick, A.S., Day, C.J., Weinstein, S.R., Griffiths, L.R., and Morrison, N.A. (2002). Differential gene expression in breast cancer cell lines and stroma-tumor differences in microdissected breast cancer biopsies revealed by display array analysis. *International journal of cancer* 100(2), 172-180.
- Mémin, E., Hoque, M., Jain, M.R., Heller, D.S., Li, H., Cracchiolo, B., et al. (2014). Blocking eIF5A modification in cervical cancer cells alters the expression of cancer-related genes and suppresses cell proliferation. *Cancer research* 74(2), 552-562.
- Merdad, A., Karim, S., Schulten, H.-J., Dallol, A., Buhmeida, A., Al-Thubaity, F., et al. (2014). Expression of matrix metalloproteinases (MMPs) in primary human breast cancer: MMP-9 as a potential biomarker for cancer invasion and metastasis. *Anticancer research* 34(3), 1355-1366.
- Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., et al. (2005). An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 102(38), 13550-13555. doi: 10.1073/pnas.0506230102.
- Minn, A.J., Gupta, G.P., Padua, D., Bos, P., Nguyen, D.X., Nuyten, D., et al. (2007). Lung metastasis genes couple breast tumor size and metastatic spread. *Proceedings of the National Academy of Sciences* 104(16), 6740-6745.
- Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., et al. (2005). Genes that mediate breast cancer metastasis to lung. *Nature* 436(7050), 518-524. doi: 10.1038/nature03799.
- Mo, D., Jiang, P., Yang, Y., Mao, X., Tan, X., Tang, X., et al. (2019). A tRNA fragment, 5'-tiRNA<sup>Val</sup>, suppresses the Wnt/ $\beta$ -catenin signaling pathway by targeting FZD3 in breast cancer. *Cancer Letters* 457, 60-73.

- Morali, O.G., Delmas, V., Moore, R., Jeanney, C., Thiery, J.P., and Larue, L.J.O. (2001). IGF-II induces rapid  $\beta$ -catenin relocation to the nucleus during epithelium to mesenchyme transition. *20(36)*, 4942-4950.
- Moreira, M.P., da Conceição Braga, L., Cassali, G.D., and Silva, L.M. (2018). STAT3 as a promising chemoresistance biomarker associated with the CD44+/high/CD24-/low/ALDH+ BCSCs-like subset of the triple-negative breast cancer (TNBC) cell line. *Experimental Cell Research* 363(2), 283-290.
- Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PloS one* 3(8), e2888.
- Munkácsy, G., Abdul-Ghani, R., Mihály, Z., Tegze, B., Tchernitsa, O., Surowiak, P., et al. (2010). PSMB7 is associated with anthracycline resistance and is a prognostic biomarker in breast cancer. *British journal of cancer* 102(2), 361-368.
- Murray, S., Briasoulis, E., Linardou, H., Bafaloukos, D., and Papadimitriou, C. (2012). Taxane resistance in breast cancer: mechanisms, predictive biomarkers and circumvention strategies. *Cancer treatment reviews* 38(7), 890-903.
- Myhre, S., Mohammed, H., Tramm, T., Alsner, J., Finak, G., Park, M., et al. (2010). In silico ascription of gene expression differences to tumor and stromal cells in a model to study impact on breast cancer outcome. *PLoS One* 5(11).
- Naoi, Y., Kishi, K., Tanei, T., Tsunashima, R., Tominaga, N., Baba, Y., et al. (2011). Development of 95-gene classifier as a powerful predictor of recurrences in node-negative and ER-positive breast cancer patients. *Breast cancer research and treatment* 128(3), 633-641.
- Neelakantan, D., Zhou, H., Oliphant, M.U., Zhang, X., Simon, L.M., Henke, D.M., et al. (2017). EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. *Nature communications* 8, 15773.
- Neubauer, H., Gall, C., Vogel, U., Hornung, R., Wallwiener, D., Solomayer, E., et al. (2008). Changes in tumour biological markers during primary systemic chemotherapy (PST). *Anticancer research* 28(3B), 1797-1804.
- Niikura, N., Liu, J., Hayashi, N., Mittendorf, E.A., Gong, Y., Palla, S.L., et al. (2012). Loss of human epidermal growth factor receptor 2 (HER2) expression in metastatic sites of HER2-overexpressing primary breast tumors. *Journal of Clinical Oncology* 30(6), 593.
- Nishizawa, Y., Nishida, N., Konno, M., Kawamoto, K., Asai, A., Koseki, J., et al. (2017). Clinical significance of histone Demethylase NO66 in invasive colorectal Cancer. *Annals of surgical oncology* 24(3), 841-849.
- O'Brien, S.K., Chen, L., Zhong, W., Armellino, D., Yu, J., Loreth, C., et al. (2015). Breast cancer cells respond differentially to modulation of TGF $\beta$ 2 signaling after exposure to chemotherapy or hypoxia. *Cancer research* 75(21), 4605-4616.
- Oguro, A., Oida, S., and Imaoka, S. (2015). Down-regulation of EPHX2 gene transcription by Sp1 under high-glucose conditions. *Biochemical Journal* 470(3), 281-291.
- Ohashi, A., Ohori, M., Iwai, K., Nambu, T., Miyamoto, M., Kawamoto, T., et al. (2015). A Novel Time-Dependent CENP-E Inhibitor with Potent Antitumor Activity. *PLoS One* 10(12), e0144675. doi: 10.1371/journal.pone.0144675.
- Okano, M., Yamamoto, H., Ohkuma, H., Kano, Y., Kim, H., Nishikawa, S., et al. (2013). Significance of INHBA expression in human colorectal cancer. *Oncology reports* 30(6), 2903-2908.
- Olkhanud, P.B., Damdinsuren, B., Bodogai, M., Gress, R.E., Sen, R., Wejksza, K., et al. (2011). Tumor-evoked regulatory B cells promote breast cancer metastasis by converting resting CD4+ T cells to T-regulatory cells. *Cancer research* 71(10), 3505-3515.
- Onder, T.T., Gupta, P.B., Mani, S.A., Yang, J., Lander, E.S., and Weinberg, R.A. (2008). Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res* 68(10), 3645-3654. doi: 10.1158/0008-5472.can-07-2938.

- Pällmann, N., Braig, M., Sievert, H., Preukschas, M., Hermans-Borgmeyer, I., Schweizer, M., et al. (2015). Biological relevance and therapeutic potential of the hypusine modification system. *Journal of Biological Chemistry* 290(30), 18343-18360.
- Pan, Z., Jing, W., He, K., Zhang, L., and Long, X. (2016). SATB1 is correlated with progression and metastasis of breast cancers: A meta-analysis. *Cellular Physiology and Biochemistry* 38(5), 1975-1983.
- Paradiso, A., Mangia, A., Chiriatti, A., Tommasi, S., Zito, A., Latorre, A., et al. (2005). Biomarkers predictive for clinical efficacy of taxol-based chemotherapy in advanced breast cancer. *Annals of Oncology* 16, iv14-iv19.
- Park, S., Kim, H., Ji, H.W., Kim, H.W., Yun, S.H., Choi, E.H., et al. (2019). Cold Atmospheric Plasma Restores Paclitaxel Sensitivity to Paclitaxel-Resistant Breast Cancer Cells by Reversing Expression of Resistance-Related Genes. *Cancers* 11(12), 2011.
- Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27(8), 1160-1167. doi: 10.1200/jco.2008.18.1370.
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* 556(7702), 463-468. doi: 10.1038/s41586-018-0040-3.
- Pavlíková, N., Bartoňová, I., Balušíková, K., Kopperova, D., Halada, P., and Kovář, J. (2015). Differentially expressed proteins in human MCF-7 breast cancer cells sensitive and resistant to paclitaxel. *Experimental cell research* 333(1), 1-10.
- Pawitan, Y., Bjohle, J., Amler, L., Borg, A.L., Eghyhazi, S., Hall, P., et al. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 7(6), R953-964. doi: 10.1186/bcr1325.
- Pernía, O., Sastre-Perona, A., Rodriguez-Antolín, C., García-Guede, A., Palomares-Bralo, M., Rosas, R., et al. (2020). A Novel Role for the Tumor Suppressor Gene ITF2 in Tumorigenesis and Chemotherapy Response. *Cancers* 12(4), 786.
- Phillips, T.M., McBride, W.H., and Pajonk, F. (2006). The response of CD24<sup>-</sup>/low/CD44<sup>+</sup> breast cancer-initiating cells to radiation. *Journal of the National Cancer Institute* 98(24), 1777-1785.
- Piasecka, D., Braun, M., Kordek, R., Sadej, R., and Romanska, H. (2018). MicroRNAs in regulation of triple-negative breast cancer progression. *Journal of cancer research and clinical oncology* 144(8), 1401-1411.
- Popiołkiewicz, J., Polkowski, K., Skierski, J.S., and Mazurek, A.P. (2005). In vitro toxicity evaluation in the development of new anticancer drugs—genistein glycosides. *Cancer Letters* 229(1), 67-75.
- Puisieux, A., Brabletz, T., and Caramel, J. (2014). Oncogenic roles of EMT-inducing transcription factors. *Nature cell biology* 16(6), 488-494.
- Pyndiah, S., Tanida, S., Ahmed, K.M., Cassimere, E.K., Choe, C., and Sakamuro, D. (2011). c-MYC suppresses BIN1 to release poly (ADP-ribose) polymerase 1: a mechanism by which cancer cells acquire cisplatin resistance. *Science signaling* 4(166), ra19-ra19.
- Raskin, L., Fullen, D.R., Giordano, T.J., Thomas, D.G., Frohm, M.L., Cha, K.B., et al. (2013). Transcriptome profiling identifies HMGA2 as a biomarker of melanoma progression and prognosis. *Journal of Investigative Dermatology* 133(11), 2585-2592.
- Reid, A.L., Millward, M., Pearce, R., Lee, M., Frank, M., Ireland, A., et al. (2013). Markers of circulating tumour cells in the peripheral blood of patients with melanoma correlate with disease recurrence and progression. *British Journal of Dermatology* 168(1), 85-92.
- Ren, J., Guo, H., Wu, H., Tian, T., Dong, D., Zhang, Y., et al. (2015). GPER in CAFs regulates hypoxia-driven breast cancer invasion in a CTGF-dependent manner. *Oncol Rep* 33(4), 1929-1937. doi: 10.3892/or.2015.3779.

- Roschke, A.V., Glebov, O.K., Lababidi, S., Gehlhaus, K.S., Weinstein, J.N., and Kirsch, I.R. (2008). Chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin maintenance. *Neoplasia (New York, NY)* 10(11), 1222.
- Roth, C., Rack, B., Müller, V., Janni, W., Pantel, K., and Schwarzenbach, H. (2010). Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. *Breast Cancer Research* 12(6), R90.
- Rouzier, R., Perou, C.M., Symmans, W.F., Ibrahim, N., Cristofanilli, M., Anderson, K., et al. (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clinical cancer research* 11(16), 5678-5685.
- Sabatier, R., Finetti, P., Adelaide, J., Guille, A., Borg, J.P., Chaffanet, M., et al. (2011). Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One* 6(11), e27656. doi: 10.1371/journal.pone.0027656.
- Sakkal, S., Miller, S., Apostolopoulos, V., and Nurgali, K. (2016). Eosinophils in cancer: favourable or unfavourable? *Current medicinal chemistry* 23(7), 650-666.
- Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. *Bioinformatics* 20(17), 3246-3248. doi: 10.1093/bioinformatics/bth349.
- Samanta, D., Gilkes, D.M., Chaturvedi, P., Xiang, L., and Semenza, G.L. (2014). Hypoxia-inducible factors are required for chemotherapy resistance of breast cancer stem cells. *Proceedings of the National Academy of Sciences* 111(50), E5429-E5438.
- Sari, E., Guler, G., Hayran, M., Gullu, I., Altundag, K., and Ozisik, Y. (2011). Comparative study of the immunohistochemical detection of hormone receptor status and HER-2 expression in primary and paired recurrent/metastatic lesions of patients with breast cancer. *Medical Oncology* 28(1), 57-63.
- Savas, P., Virassamy, B., Ye, C., Salim, A., Minto, C.P., Caramia, F., et al. (2018). Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nature medicine* 24(7), 986-993.
- Saxena, M., Stephens, M.A., Pathak, H., and Rangarajan, A. (2011). Transcription factors that mediate epithelial-mesenchymal transition lead to multidrug resistance by upregulating ABC transporters. *Cell death & disease* 2(7), e179-e179.
- Schmalhofer, O., Brabletz, S., Brabletz, T.J.C., and Reviews, M. (2009). E-cadherin,  $\beta$ -catenin, and ZEB1 in malignant progression of cancer. 28(1-2), 151-166.
- Schmidt, M., Böhm, D., von Törne, C., Steiner, E., Puhl, A., Pilch, H., et al. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 68(13), 5405-5413. doi: 10.1158/0008-5472.can-07-5206.
- Scully, O.J., Bay, B.-H., Yip, G., and Yu, Y. (2012). Breast cancer metastasis. *Cancer Genomics-Proteomics* 9(5), 311-320.
- Seder, C.W., Hartojo, W., Lin, L., Silvers, A.L., Wang, Z., Thomas, D.G., et al. (2009). Upregulated INHBA expression may promote cell proliferation and is associated with poor survival in lung adenocarcinoma. *Neoplasia* 11(4), 388-396.
- Senses, K.M., Ghasemi, M., Akbar, M.W., Isbilen, M., Fallacara, A.L., Frankenburg, S., et al. (2017). Phenotype-based variation as a biomarker of sensitivity to molecularly targeted therapy in melanoma. *Medchemcomm* 8(1), 88-95. doi: 10.1039/c6md00466k.
- Shafee, N., Smith, C.R., Wei, S., Kim, Y., Mills, G.B., Hortobagyi, G.N., et al. (2008). Cancer stem cells contribute to cisplatin resistance in Brca1/p53-mediated mouse mammary tumors. *Cancer research* 68(9), 3243-3250.
- Shang, Y., Cai, X., and Fan, D. (2013). Roles of epithelial-mesenchymal transition in cancer drug resistance. *Curr Cancer Drug Targets* 13(9), 915-929.
- Shats, I., Gatz, M.L., Chang, J.T., Mori, S., Wang, J., Rich, J., et al. (2011). Using a stem cell-based signature to guide therapeutic selection in cancer. *Cancer research* 71(5), 1772-1780.

- Shaw, F.L., Harrison, H., Spence, K., Ablett, M.P., Simões, B.M., Farnie, G., et al. (2012). A detailed mammosphere assay protocol for the quantification of breast stem cell activity. *Journal of mammary gland biology and neoplasia* 17(2), 111-117.
- Shekhar, M.P., Lyakhovich, A., Visscher, D.W., Heng, H., and Kondrat, N. (2002). Rad6 overexpression induces multinucleation, centrosome amplification, abnormal mitosis, aneuploidy, and transformation. *Cancer Res* 62(7), 2115-2124.
- Shi, X., Zhan, L., Xiao, C., Lei, Z., Yang, H., Wang, L., et al. (2015). miR-1238 inhibits cell proliferation by targeting LHX2 in non-small cell lung cancer. *Oncotarget* 6(22), 19043.
- Shriver, M., Stroka, K., Vitolo, M., Martin, S., Huso, D., Konstantopoulos, K., et al. (2015). Loss of giant obscurins from breast epithelium promotes epithelial-to-mesenchymal transition, tumorigenicity and metastasis. *Oncogene* 34(32), 4248-4259.
- Siegel, M.B., He, X., Hoadley, K.A., Hoyle, A., Pearce, J.B., Garrett, A.L., et al. (2018). Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer. *The Journal of clinical investigation* 128(4), 1371-1383.
- Siegel, R.L., Miller, K.D., and Jemal, A.J.C.a.c.j.f.c. (2019). Cancer statistics, 2019. 69(1), 7-34.
- Sievert, H., Pällmann, N., Miller, K.K., Hermans-Borgmeyer, I., Venz, S., Sendoel, A., et al. (2014). A novel mouse model for inhibition of DOHH-mediated hypusine modification reveals a crucial function in embryonic development, proliferation and oncogenic transformation. *Disease models & mechanisms* 7(8), 963-976.
- Silver, D.P., Richardson, A.L., Eklund, A.C., Wang, Z.C., Szallasi, Z., Li, Q., et al. (2010). Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. *Journal of clinical oncology* 28(7), 1145.
- Simpson, C.D., Anyiwe, K., and Schimmer, A.D. (2008). Anoikis resistance and tumor metastasis. *Cancer letters* 272(2), 177-185.
- Sinnberg, T., Levesque, M.P., Krochmann, J., Cheng, P.F., Ikenberg, K., Meraz-Torres, F., et al. (2018). Wnt-signaling enhances neural crest migration of melanoma cells and induces an invasive phenotype. *Molecular cancer* 17(1), 59.
- Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A.M., Yu, J., Klijn, J.G., et al. (2008). Subtypes of breast cancer show preferential site of relapse. *Cancer research* 68(9), 3108-3114.
- Smith, L., Watson, M.B., O'Kane, S.L., Drew, P.J., Lind, M.J., and Cawkwell, L. (2006). The analysis of doxorubicin resistance in human breast cancer cells using antibody microarrays. *Molecular cancer therapeutics* 5(8), 2115-2120.
- Somasagara, R.R., Spencer, S.M., Tripathi, K., Clark, D.W., Mani, C., Madeira da Silva, L., et al. (2017). RAD6 promotes DNA repair and stem cell signaling in ovarian cancer and is a promising therapeutic target to prevent and treat acquired chemoresistance. *Oncogene* 36(48), 6680-6690. doi: 10.1038/onc.2017.279.
- Stockwin, L.H., Sherry, X.Y., Stotler, H., Hollingshead, M.G., and Newton, D.L. (2009). ARC (NSC 188491) has identical activity to Sangivamycin (NSC 65346) including inhibition of both P-TEFb and PKC. *BMC cancer* 9(1), 1-13.
- Su, S., Liao, J., Liu, J., Huang, D., He, C., Chen, F., et al. (2017). Blocking the recruitment of naive CD4<sup>+</sup> T cells reverses immunosuppression in breast cancer. *Cell research* 27(4), 461-482.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43), 15545-15550. doi: 10.1073/pnas.0506580102.
- Sun, F., Wang, P., Luan, R., Zou, S., and Du, X. (2018). MicroRNA-574 enhances doxorubicin resistance through down-regulating SMAD4 in breast cancer cells. *Eur Rev Med Pharmacol Sci* 22(5), 1342-1350.

- Sun, S., Liang, X., Zhang, X., Liu, T., Shi, Q., Song, Y., et al. (2015). Phosphoglycerate kinase-1 is a predictor of poor survival and a novel prognostic biomarker of chemoresistance to paclitaxel treatment in breast cancer. *British journal of cancer* 112(8), 1332-1339.
- Sun, S., Sun, Y., Rong, X., and Bai, L. (2019). High glucose promotes breast cancer proliferation and metastasis by impairing angiotensinogen expression. *Bioscience reports* 39(6).
- Suzuki, C., Takahashi, K., Hayama, S., Ishikawa, N., Kato, T., Ito, T., et al. (2007). Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer. *Molecular cancer therapeutics* 6(2), 542-551.
- Szved, M., Kania, K.D., and Jozwiak, Z. (2015). Toxicity of doxorubicin-transferrin conjugate is connected to the modulation of Wnt/ $\beta$ -catenin pathway in human leukemia cells. *Leukemia research* 39(10), 1096-1102.
- Tacca, O., Penault-Llorca, F., Abrial, C., Mouret-Reynier, M.-A., Raoelfils, I., Durando, X., et al. (2007). Changes in and prognostic value of hormone receptor status in a series of operable breast cancer patients treated with neoadjuvant chemotherapy. *Oncologist* 12(6), 636.
- Takebe, N., Warren, R.Q., and Ivy, S.P. (2011). Breast cancer growth and metastasis: interplay between cancer stem cells, embryonic signaling pathways and epithelial-to-mesenchymal transition. *Breast cancer research* 13(3), 1-11.
- Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y., et al. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med* 6(10), 1279-1293. doi: 10.15252/emmm.201404208.
- Tanaka, H., Kamita, S.G., Wolf, N.M., Harris, T.R., Wu, Z., Morisseau, C., et al. (2008). Transcriptional regulation of the human soluble epoxide hydrolase gene EPHX2. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1779(1), 17-27.
- Taube, J.H., Herschkowitz, J.I., Komurov, K., Zhou, A.Y., Gupta, S., Yang, J., et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A* 107(35), 15449-15454. doi: 10.1073/pnas.1004900107.
- Team, R.C. (2013). "R: A language and environment for statistical computing". Vienna, Austria.
- Therneau, T.M., and Grambsch, P.M. (2000). "The Cox model," in *Modeling survival data: extending the Cox model*. Springer), 39-77.
- Thomassen, M., Tan, Q., and Kruse, T.A. (2009). Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast cancer research and treatment* 113(2), 239-249.
- Thompson, A.M., Jordan, L.B., Quinlan, P., Anderson, E., Skene, A., Dewar, J.A., et al. (2010). Prospective comparison of switches in biomarker status between primary and recurrent breast cancer: the Breast Recurrence In Tissues Study (BRITS). *Breast cancer research* 12(6), R92.
- Tiebe, M., Lutz, M., Tiebe, D.S., and Teleman, A.A. (2019). Crebl2 regulates cell metabolism in muscle and liver cells. *Scientific Reports* 9(1), 1-12.
- Tomida, S., Yanagisawa, K., Koshikawa, K., Yatabe, Y., Mitsudomi, T., Osada, H., et al. (2007). Identification of a metastasis signature and the DLX4 homeobox protein as a regulator of metastasis by combined transcriptome approach. *Oncogene* 26(31), 4600-4608. doi: 10.1038/sj.onc.1210242.
- Tseng, Y.-J., Huang, C.-E., Wen, C.-N., Lai, P.-Y., Wu, M.-H., Sun, Y.-C., et al. (2019). Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. *International journal of medical informatics* 128, 79-86.
- Valent, P., Bonnet, D., De Maria, R., Lapidot, T., Copland, M., Melo, J.V., et al. (2012). Cancer stem cell definitions and terminology: the devil is in the details. *Nature Reviews Cancer* 12(11), 767-775.

- van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871), 530-536. doi: 10.1038/415530a.
- Van Phuc, P., Nhan, P.L.C., Nhung, T.H., Tam, N.T., Hoang, N.M., Tue, V.G., et al. (2011). Downregulation of CD44 reduces doxorubicin resistance of CD44<sup>+</sup> CD24<sup>-</sup> breast cancer cells. *OncoTargets and therapy* 4, 71.
- Vargas, G., Bouchet, M., Bouazza, L., Reboul, P., Boyault, C., Gervais, M., et al. (2019). ER $\alpha$  promotes breast cancer cell dissemination to bone by increasing RANK expression in primary breast tumors. *Oncogene* 38(7), 950-964.
- Velasco-Velázquez, M.A., Homsí, N., De La Fuente, M., and Pestell, R.G. (2012). Breast cancer stem cells. *The international journal of biochemistry & cell biology* 44(4), 573-577.
- Vera-Ramirez, L., Sanchez-Rovira, P., Ramirez-Tortosa, C.L., Quiles, J.L., Ramirez-Tortosa, M., and Lorente, J.A. (2013). Transcriptional shift identifies a set of genes driving breast cancer chemoresistance. *PLoS One* 8(1), e53983. doi: 10.1371/journal.pone.0053983.
- Vieira, A.F., Ricardo, S., Ablett, M.P., Dionísio, M.R., Mendes, N., Albergaria, A., et al. (2012). P-cadherin is coexpressed with CD44 and CD49f and mediates stem cell properties in basal-like breast cancer. *Stem cells* 30(5), 854-864.
- Vinci, M., Gowan, S., Boxall, F., Patterson, L., Zimmermann, M., Lomas, C., et al. (2012). Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. *BMC biology* 10(1), 29.
- Vogel, C., Malter, W., Morgenstern, B., Ludwig, S., Vehreschild, J.J., Hamacher, S., et al. (2019). The Role of Previous Therapies and Sites of Metastasis as Influencing Factors on Discordance of ER, PR and HER2 Status Between Primary and Metastasized Breast Cancer. *Anticancer research* 39(5), 2647-2659.
- Voller, D., Reinders, J., Meister, G., and Bosserhoff, A.K. (2013). Strong reduction of AGO2 expression in melanoma and cellular consequences. *Br J Cancer* 109(12), 3116-3124. doi: 10.1038/bjc.2013.646.
- Wang, L., Tan, C., Qiao, F., Wang, W., Jiang, X., Lian, P., et al. (2015). Upregulated expression of DIXDC1 in intestinal-type gastric carcinoma: co-localization with beta-catenin and correlation with poor prognosis. *Cancer Cell Int* 15, 120. doi: 10.1186/s12935-015-0273-2.
- Wang, R., Lv, Q., Meng, W., Tan, Q., Zhang, S., Mo, X., et al. (2014). Comparison of mammosphere formation from breast cancer cell lines and primary breast tumors. *Journal of thoracic disease* 6(6), 829.
- Wang, X., Wang, G., Zhao, Y., Liu, X., Ding, Q., Shi, J., et al. (2012). STAT3 mediates resistance of CD44<sup>+</sup> CD24<sup>-</sup>/low breast cancer stem cells to tamoxifen in vitro. *Journal of biomedical research* 26(5), 325-335.
- Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365(9460), 671-679. doi: 10.1016/s0140-6736(05)17947-1.
- Watson, M.A., Darrow, C., Zimonjic, D.B., Popescu, N.C., and Fleming, T.P. (1998). Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13. *Oncogene* 16(6), 817-824.
- Weng, C.H., Chen, L.Y., Lin, Y.C., Shih, J.Y., Lin, Y.C., Tseng, R.Y., et al. (2019). Epithelial-mesenchymal transition (EMT) beyond EGFR mutations per se is a common mechanism for acquired resistance to EGFR TKI. *Oncogene* 38(4), 455-468. doi: 10.1038/s41388-018-0454-2.
- Widmer, D.S., Hoek, K.S., Cheng, P.F., Eichhoff, O.M., Biedermann, T., Raaijmakers, M.I., et al. (2013). Hypoxia contributes to melanoma heterogeneity by triggering HIF1 $\alpha$ -dependent phenotype switching. *Journal of Investigative Dermatology* 133(10), 2436-2443.

- Xie, Q., Wang, S., Zhao, Y., Zhang, Z., Qin, C., and Yang, X. (2017). MiR-519d impedes cisplatin-resistance in breast cancer stem cells by down-regulating the expression of MCL-1. *Oncotarget* 8(13), 22003.
- Xie, S., Ogden, A., Aneja, R., and Zhou, J. (2016). Microtubule-binding proteins as promising biomarkers of paclitaxel sensitivity in cancer chemotherapy. *Medicinal research reviews* 36(2), 300-312.
- Xu, H., Dephoure, N., Sun, H., Zhang, H., Fan, F., Liu, J., et al. (2015). Proteomic profiling of paclitaxel treated cells identifies a novel mechanism of drug resistance mediated by PDCD4. *Journal of Proteome Research* 14(6), 2480-2491.
- Xu, J., Sun, T., Guo, X., Wang, Y., and Jing, M. (2018a). Estrogen receptor- $\alpha$  promoter methylation is a biomarker for outcome prediction of cisplatin resistance in triple-negative breast cancer. *Oncology letters* 15(3), 2855-2862.
- Xu, L., Cheng, L., Yang, F., Pei, B., Liu, X., Zhou, J., et al. (2018b). JWA suppresses the invasion of human breast carcinoma cells by downregulating the expression of CXCR4. *Molecular medicine reports* 17(6), 8137-8144.
- Yamada, K.H., Nakajima, Y., Geyer, M., Wary, K.K., Ushio-Fukai, M., Komarova, Y., et al. (2014). KIF13B regulates angiogenesis through Golgi to plasma membrane trafficking of VEGFR2. *J Cell Sci* 127(20), 4518-4530.
- Yan, W.-T., Cui, X., Chen, Q., Li, Y.-F., Cui, Y.-H., Wang, Y., et al. (2017). Circulating tumor cell status monitors the treatment responses in breast cancer patients: a meta-analysis. *Scientific reports* 7, 43464.
- Yan, W., Cao, Q.J., Arenas, R.B., Bentley, B., and Shao, R. (2010). GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *Journal of Biological Chemistry* 285(18), 14042-14051.
- Yang, F., Teves, S.S., Kemp, C.J., and Henikoff, S. (2014). Doxorubicin, DNA torsion, and chromatin dynamics. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1845(1), 84-89.
- Yang, L., Lin, C., and Liu, Z.-R.J.C. (2006). P68 RNA helicase mediates PDGF-induced epithelial mesenchymal transition by displacing Axin from  $\beta$ -catenin. *127(1)*, 139-155.
- Yang, Z., Li, N., Li, X., Lei, L., and Wang, X. (2020). The Prognostic Impact of Hormonal Receptor and HER-2 Expression Discordance in Metastatic Breast Cancer Patients. *OncoTargets and therapy* 13, 853.
- Yardimci, H., van Duffelen, M., Mao, Y., Rosenfeld, S.S., and Selvin, P.R. (2008). The mitotic kinesin CENP-E is a processive transport motor. *Proc Natl Acad Sci U S A* 105(16), 6016-6021. doi: 10.1073/pnas.0711314105.
- Ye, L., Guo, L., He, Z., Wang, X., Lin, C., Zhang, X., et al. (2016). Upregulation of E2F8 promotes cell proliferation and tumorigenicity in breast cancer by modulating G1/S phase transition. *Oncotarget* 7(17), 23757-23771. doi: 10.18632/oncotarget.8121.
- Yeong, J., Lim, J.C.T., Lee, B., Li, H., Chia, N., Ong, C.C.H., et al. (2018). High densities of tumor-associated plasma cells predict improved prognosis in triple negative breast cancer. *Frontiers in immunology* 9, 1209.
- Yu, M., Gong, J., Ma, M., Yang, H., Lai, J., Wu, H., et al. (2009a). Immunohistochemical analysis of human arrest-defective-1 expressed in cancers in vivo. *Oncol Rep* 21(4), 909-915. doi: 10.3892/or\_00000303.
- Yu, M., Ma, M., Huang, C., Yang, H., Lai, J., Yan, S., et al. (2009b). Correlation of expression of human arrest-defective-1 (hARD1) protein with breast cancer. *Cancer Invest* 27(10), 978-983. doi: 10.3109/07357900902769723.
- Zahran, A.M., Shaltout, A.S., Fakhry, H., Abdel Fattah, O.N., Temerik, D.F., Khalaf, S.M., et al. (2020). Prognostic Significance of Circulating CD28 Negative Suppressor T Cells and Memory B Cells in Patients with Breast Cancer. *Iranian Journal of Immunology* 17(2), 95-110.

- Zang, S., Chen, F., Dai, J., Guo, D., Tse, W., Qu, X., et al. (2010). RNAi-mediated knockdown of Notch-1 leads to cell growth inhibition and enhanced chemosensitivity in human breast cancer. *Oncology reports* 23(4), 893-899.
- Zeng, D., Liang, Y.K., Xiao, Y.S., Wei, X.L., Lin, H.Y., Wu, Y., et al. (2020). Inhibition of Notch1 reverses EMT and chemoresistance to cisplatin via direct downregulation of MCAM in triple-negative breast cancer cells. *International Journal of Cancer*.
- Zeng, Y., Min, L., Han, Y., Meng, L., Liu, C., Xie, Y., et al. (2014). Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis. *Carcinogenesis* 35(10), 2244-2253. doi: 10.1093/carcin/bgu132.
- Zhai, Q., Li, H., Sun, L., Yuan, Y., and Wang, X. (2019). Identification of differentially expressed genes between triple and non-triple-negative breast cancer using bioinformatics analysis. *Breast Cancer* 26(6), 784-791.
- Zhang, H., Wang, Y., Dou, J., Guo, Y., He, J., Li, L., et al. (2019). Acetylation of AGO2 promotes cancer progression by increasing oncogenic miR-19b biogenesis. *Oncogene* 38(9), 1410-1431. doi: 10.1038/s41388-018-0530-7.
- Zhang, Z., Ni, C., Chen, W., Wu, P., Wang, Z., Yin, J., et al. (2014). Expression of CXCR4 and breast cancer prognosis: a systematic review and meta-analysis. *BMC cancer* 14(1), 49.
- Zhao, L., Ge, C., Zhang, Z., Hu, H., Zhang, Y., Zhao, W., et al. (2020). FAM136A immunoreactivity is associated with nodal involvement and survival in lung adenocarcinoma in a Chinese case series. *Bioengineered* 11(1), 261-271.
- Zhao, W., Li, Y., and Zhang, X. (2017). Stemness-related markers in cancer. *Cancer translational medicine* 3(3), 87.
- Zhao, Y., Hu, J., Miao, G., Qu, L., Wang, Z., Li, G., et al. (2013). Transmembrane protein 208: a novel ER-localized protein that regulates autophagy and ER stress. *PLoS One* 8(5), e64228. doi: 10.1371/journal.pone.0064228.
- Zhou, S., Shen, J., Lin, S., Liu, X., Xu, M., Shi, L., et al. (2016). Downregulated expression of DIXDC1 in hepatocellular carcinoma and its correlation with prognosis. *Tumour Biol* 37(10), 13607-13616. doi: 10.1007/s13277-016-5213-9.
- Zhu, L., Zou, J., Zhao, Y., Jiang, X., Wang, Y., Wang, X., et al. (2018). ER- $\alpha$ 36 mediates cisplatin resistance in breast cancer cells through EGFR/HER-2/ERK signaling pathway. *Journal of Experimental & Clinical Cancer Research* 37(1), 123.
- Zidan, J., Dashkovsky, I., Stayerman, C., Basher, W., Cozacov, C., and Hadary, A. (2005). Comparison of HER-2 overexpression in primary breast cancer and metastatic sites and its effect on biological targeting therapy of metastatic disease. *British journal of cancer* 93(5), 552-556.
- Zingg, D., Debbache, J., Schaefer, S.M., Tuncer, E., Frommel, S.C., Cheng, P., et al. (2015). The epigenetic modifier EZH2 controls melanoma growth and metastasis through silencing of distinct tumour suppressors. *Nature communications* 6(1), 1-17.
- Zou, J.X., Duan, Z., Wang, J., Sokolov, A., Xu, J., Chen, C.Z., et al. (2014). Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. *Molecular Cancer Research* 12(4), 539-549.
- Zurita, M., Lara, P.C., del Moral, R., Torres, B., Linares-Fernández, J.L., Arrabal, S.R., et al. (2010). Hypermethylated 14-3-3- $\sigma$  and ESR1 gene promoters in serum as candidate biomarkers for the diagnosis and treatment efficacy of breast cancer metastasis. 10(1), 217.

## 9. Appendix

**Table 9.1: Datasets used for validation of CNCL gene list.**

| Dataset                 | Original findings                                                                                                                                                                                                                                                                                                    | Findings of this study                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>GSE36139 (CCLE)</b>  | Authors developed and analyzed a data resource which contained gene expression, copy number and drug cytotoxicity data for 947 cell lines. They showed that cell lines do represent subtypes of various cancer and drug response data generated here could help in development of personalized therapeutic regimens. | We used 56 breast cancer cell lines' gene expression data to find differentially expressed genes between CS/M and NS/E groups. Drug cytotoxicity data was used to identify drugs which could target discovered groups separately.                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>E-MTAB-783 (CGP)</b> | Cell lines were screened with 130 different drugs.                                                                                                                                                                                                                                                                   | We used 39 breast cancer cell lines' gene expression data to find differentially expressed genes between CS/M and NS/E groups. Drug cytotoxicity data was used to identify drugs which could target discovered groups separately.                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>GSE24717</b>         | Authors developed a stemness signature to differentiate between cancer stem cell enriched samples. This signature has prognostic importance and authors recommend to treat stem cell enriched samples with topoisomerase inhibitors and resveratrol.                                                                 | To identify this signature, authors did not used breast cancer specific cancer stem cell markers (CD44/CD24) but used CD133. But they later showed if they classify stem enriched cell lines from the rest then both markers show same pattern. But on the other hand our signature can identify not only CS/M group (CD44+/CD24-) from NS/E group (CD44-/CD24+) but can also differentiate between resistant and sensitive cell lines to several commercial drugs and also our classification overlaps with epithelial and mesenchymal classification as evidenced in literature as well. We used this dataset to show that using our |

|                 |                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 |                                                                                                                                                                                                                                                                                                                                                              | differentially expressed signature can classify breast cancer cell lines in the same groups which were formed in our discovery dataset, CCLE and CGP.                                                                                                                                                                                                                                                    |
| <b>GSE50811</b> | Authors performed gene expression profiling of breast cancer cell lines and used this data to identify genes which can be related with paclitaxel and eribulin sensitivity. They showed that EMT genes were related eribulin sensitivity.                                                                                                                    | In this paper authors treated cell lines with paclitaxel and eribulin only for 24 hours before checking their gene expression. This time is not enough for such experiment. So we used only untreated cell line data to further validate our gene signature in clustering breast cancer cell lines. Cell lines were clustered into same groups which were formed in our discovery dataset, CCLE and CGP. |
| <b>GSE73526</b> | Authors performed shRNA dropout screens on 77 breast cancer cell lines to identify vulnerabilities in breast cancer and associated this data with genomic and proteomic data of those cell lines. Additionally comparing those vulnerabilities with drug data showed potential resistance mechanisms, anticancer effects and need for combination therapies. | We used this dataset to show that using our gene signature can classify breast cancer cell lines in the same groups which were formed via our discovery datasets, CCLE and CGP.                                                                                                                                                                                                                          |
| <b>GSE15192</b> | Authors showed that a subpopulation of MCF-10A cells acquire CD44+/CD24- phenotype, and that a few EMT related genes play a role in this switch. They found 2035 genes as differentially expressed, and validated some.                                                                                                                                      | We developed a gene signature that can differentiate between CD44+/CD24- and CD44-/CD24+ phenotypes. And we used gene expression data uploaded by the authors to validate this signature and successfully clustered samples as expected.                                                                                                                                                                 |

|                 |                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                     |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>GSE36643</b> | Authors investigated a new CSC marker GD2, in HMLER cells and proposed to use this as a single marker of CSC as opposed to CD44 and CD24 markers for breast cancer.                                                                                                                                                                       | We utilized CD44 and CD24 based distinctions to validate our gene list successfully.                                                                                                |
| <b>GSE52327</b> | Authors sorted patient derived breast cancer cells based on ALDH, another marker for stemness.                                                                                                                                                                                                                                            | We showed that CD44 and CD24 gene expression does not correlate with ALDH gene expression. We used this dataset to validate this observation.                                       |
| <b>GSE9691</b>  | Authors investigated the role of E-cadherin loss in promoting metastasis and concluded that its loss in breast cancer HMLE cells not only increases their metastatic potential but also increases their invasiveness, motility and resistance to apoptosis.                                                                               | We could identify E-cad downregulated samples as CS/M from control and beta catenin downregulated samples as NS/E.                                                                  |
| <b>GSE24202</b> | Authors used this dataset to associate EMT with breast cancer stem cells. They generated mesenchymal cells HMLE cells by overexpressing TGF beta, Twist, Gsc and by downregulating E-cad. They identified a gene signature of 159 transcription factors responsible for clustering mesenchymal/stem cells from Epithelial/non stem cells. | We used this dataset to successfully distinguish epithelial and mesenchymal cell groups generated by the authors with the exception of siE-cad cells, which we explain in figure 3. |
| <b>GSE7515</b>  | Mammosphere culture is associated with enriching cells for cancer stem cells. Authors generated this dataset from human breast tumor cells cultured in adherent conditions and mammosphere culture. Their aim was to identify genes which                                                                                                 | We used this datasets to further validate CNCL and most of mammospheres were clustered as CS/M and primary breast cancer cell lines as NS/E.                                        |

|                 |                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                        |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 | could distinguish adherent cells from mammospheres.                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                        |
| <b>GSE24460</b> | Authors generated doxorubicin resistant MCF7 cells which were highly invasive, tumorigenic and formed mammospheres when compared to control cells. 30% of these MCF7 doxorubicin resistant cells showed CD44+/CD24- phenotype. Genes responsible for drug resistance and stem cell characteristics were high in resistant cells when compared to sensitive cells. | We used this dataset to find if CNCL can identify resistant MCF7 cells from controls. Upon hierarchical clustering, as expected resistant cells were clustered as CS/M separately from control cells as NS/E.                                          |
| <b>GSE10281</b> | Stem cells are responsible for drug resistance. Authors took biopsies from patients before treatment and after treatment with letrozole for 3 months. They looked at the mesenchymal and epithelial markers and these were differentially expressed in samples before and after letrozole therapy.                                                                | We used this dataset to show that CNCL can identify patients before and after undergoing treatment. Half of NS/E samples switched to a CS/M phenotype and only one patient switched in the opposite direction while others maintained their phenotype. |
| <b>GSE12791</b> | In this dataset, authors developed paclitaxel resistance in breast cancer cell line MDAMB231 by prolonged drug treatment and studied the effect of Bexarotene in switching resistant phenotype back to sensitive.                                                                                                                                                 | We used this dataset to successfully cluster paclitaxel resistant phenotype (CS/M) from sensitive phenotype (NS/E).                                                                                                                                    |
| <b>GSE23399</b> | Breast cancer associated fibroblasts (CAF) were isolated from patients' tumor specimens and were treated with paclitaxel over a prolonged time. These chemotherapy resistant CAFs are                                                                                                                                                                             | We used this datasets to successfully demonstrate that drug resistant phenotype behaves as CS/M and control cells behave like NS/E cells.                                                                                                              |

|                 |                                                                                                                                                                                 |                                                                                                                                                                                   |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 | responsible for tumor growth and aggression.                                                                                                                                    |                                                                                                                                                                                   |
| <b>GSE16179</b> | Authors treated breast cancer cell line BT474 with lapatinib over a prolonged period of time and demonstrated that AXL plays a novel role in acquiring resistance to Lapatinib. | We used Lapatinib sensitive and resistant cell models to successfully demonstrate that the resistant phenotype is of a CS/M, while the sensitive phenotype is classified as NS/E. |
| <b>GSE28844</b> | In this study authors aimed to identify such pathways which confer resistance to tumors post chemotherapy.                                                                      | We used this dataset to show that tumors treated with Taxane have a higher CS/M score when compared to pretreated samples                                                         |

**Table 9.2: Datasets used in prognostic evaluation for CNCL gene list**

| <b>Dataset</b> | <b>Original findings</b>                                                                                                                     | <b>Findings of this study</b>                                                                                                                                                                                                         |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>GSE1456</b> | Authors developed a 64 gene signature which can estimate breast cancer patients' response to adjuvant therapy.                               | Our survival analysis using CNCL revealed that patients with NS/E phenotype showed worse prognosis significantly when compared with CS/M phenotype, using disease specific survival, Overall survival and relapse free survival data. |
| <b>GSE2034</b> | Authors developed a 76 gene signature which can identify patients at high risk of distant recurrence from patients with favorable prognosis. | CNCL showed no difference in recurrence between CS/M and NS/E patients.                                                                                                                                                               |
| <b>GSE2603</b> | Authors identified genes which are responsible for breast cancer metastasis to bone and lung tissue.                                         | CNCL showed that patients with CS/M phenotype had worse prognosis when compared with NS/E patients for lymph node metastasis free survival.                                                                                           |
| <b>GSE3494</b> | Authors identified a 32 gene signature which can differentiate between p53 wild type and mutant samples, and predicts                        | CNCL showed no significant difference when patients with CS/M phenotype were compared with NS/E patients.                                                                                                                             |

|                 |                                                                                                                                                                    |                                                                                                                                          |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
|                 | survival independent of all other prognostic factors.                                                                                                              |                                                                                                                                          |
| <b>GSE4922</b>  | Authors identified a genetic grade signature which can separate low and high grade disease and can improve therapeutic decision making for breast cancer patients. | CNCL showed patients with CS/M phenotype showed better prognosis when compared with NS/E patients with border line significance.         |
| <b>GSE6532</b>  | Authors developed a gene grade index which defined histologic grade and found 2 distinct ER+ subgroups with survival difference.                                   | CNCL showed patients with CS/M phenotype showed significantly better prognosis when compared with NS/E patients.                         |
| <b>GSE7390</b>  | Authors validated a 76 gene signature for distant metastasis free survival, overall survival, relapse free survival, time to distant metastasis survival.          | CNCL showed no survival difference between CS/M and NS/E patients.                                                                       |
| <b>GSE11121</b> | Authors generated and associated several metagenes with distant metastasis free survival (proliferation metagene and B cell metagene)                              | CNCL showed patients with CS/M phenotype showed better prognosis when compared with NS/E patients which was statistically insignificant. |
| <b>GSE12276</b> | Authors identified genes which are responsible for breast cancer metastasis to brain (COX2, HBEGF and ST6GALNAC5).                                                 | CNCL showed no survival difference between CS/M and NS/E patients.                                                                       |
| <b>GSE19615</b> | Authors identified 2 genes (LAPTM4B and YWHAZ) as responsible for generation of chemoresistance to anthracyclines.                                                 | CNCL showed no survival difference between CS/M and NS/E patients.                                                                       |
| <b>GSE20685</b> | Authors identified molecular subtypes of breast cancer and proposed these subtypes to better customization of breast cancer treatment.                             | CNCL showed no survival difference between CS/M and NS/E patients.                                                                       |

|                                 |                                                                                                                                                                       |                                                                                                                                                                                        |
|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>GSE21653</b>                 | Authors suggested ECRG4 as tumor suppressor gene which can be used to better breast cancer prognostication.                                                           | CNCL showed no survival difference between CS/M and NS/E patients.                                                                                                                     |
| <b>GSE58812</b>                 | Authors identified 3 subtypes of triple negative breast cancer and proposed that immune mediation in these tumors can be channeled to treat specific subtypes.        | CNCL showed patients with CS/M phenotype showed better prognosis when compared with NS/E patients significantly for metastasis free survival and insignificantly for overall survival. |
| <b>GSE25066</b>                 | Authors developed a genomic predictor for patients treated with taxane and anthracycline chemotherapy.                                                                | CNCL showed patients with CS/M phenotype showed worse prognosis when compared with NS/E patients with statistical significance.                                                        |
| <b>Metabric British Cohort</b>  | Authors performed unsupervised analysis of paired DNA RAN profiles and found novel groups with distinct clinical outcomes and then validated these in another cohort. | CNCL showed patients with CS/M phenotype showed worse prognosis when compared with NS/E patients with statistical significance.                                                        |
| <b>Metabric Canadian Cohort</b> |                                                                                                                                                                       | CNCL showed no survival difference between CS/M and NS/E patients.                                                                                                                     |

## 10. Publications

### Copyright Permissions

#### 1: Journal of Cancer

“A Stemness and EMT Based Gene Expression Signature Identifies Phenotypic Plasticity and is A Predictive but Not Prognostic Biomarker for Breast Cancer”

##### Permission request

**Request to reproduce figures or tables in other publications or journals:** There is no need to obtain permission if copyrights of the figures or tables belong to Ivyspring International Publisher. Please acknowledge the original source of publication in our journal with full citation. If the copyrights of the figures or tables belong to other publishers or authors, please contact the original publishers or authors for permissions.

**Authors or co-authors:** As an author or co-author of the article, you are free to include or reuse any part of the materials in your thesis or other publications that you write, with citation of the original source in our journal. There is no need to obtain permission. (See also our [terms of use](#)).

#### 2: MedChemComm

“Phenotype-based variation as a biomarker of sensitivity to molecularly targeted therapy in melanoma”

### Re-use permission requests

Material published by the Royal Society of Chemistry and other publishers is subject to all applicable copyright, database protection, and other rights. Therefore, for any publication, whether printed or electronic, permission must be obtained to use material for which the author(s) does not already own the copyright. This material may be, for example, a figure, diagram, table, photo or some other image.

#### Author reusing their own work published by the Royal Society of Chemistry

You do not need to request permission to reuse your own figures, diagrams, etc, that were originally published in a Royal Society of Chemistry publication. However, permission should be requested for use of the whole article or chapter except if reusing it in a thesis. If you are including an article or book chapter published by us in your thesis please ensure that your co-authors are aware of this.

Reuse of material that was published originally by the Royal Society of Chemistry must be accompanied by the appropriate acknowledgement of the publication. The form of the acknowledgement is dependent on the journal in which it was published originally, as detailed in 'Acknowledgements'.

#### 3: ACS Medicinal Chemistry Letters

“Predictive Gene Signature for Pyrazolopyrimidine Derivative c-Src Inhibitor 10a Sensitivity in Melanoma Cells”

**Predictive Gene Signature for Pyrazolopyrimidine Derivative c-Src Inhibitor 10a Sensitivity in Melanoma Cells**

Author: Baris Kucukkaraduman, Can Turk, Anna L. Fallacara, et al

Publication: ACS Medicinal Chemistry Letters

Publisher: American Chemical Society

Date: May 1, 2020

Copyright © 2020, American Chemical Society

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)[CLOSE WINDOW](#)

#### 4: BIORXIV

### “Plasticity as an Underlying Mechanism of Tumor Heterogeneity in Breast Cancer”



**Bilkent University**

Department of Molecular Biology and Genetics

12 September 2020

To whom it may concern:

I allow Muhammad Waqas Akbar to use the material in the following paper in his thesis with the citation of the original source.

Paper:

“M. W. Akbar., M. Isbilen, B. Kucukkaraduman, S. D. Canli, E. Dedeoglu, S. Azizolli, I. Cela, A. G. Akcay, H. Hakanoglu and A. O. Gure (2020). "Plasticity as an Underlying Mechanism of Tumor Heterogeneity in Breast Cancer." BIORXV. 2020.2003.2011.987511.”

Best regards,

Ali Osmay Gure MD, PhD

Chair,  
Department of Molecular Biology and Genetics  
Bilkent University, SB-248,  
Cankaya, Ankara

+90 (312) 290 2507 - tel.  
+90 (312) 266 5097 - facs.  
+90 (533) 230 7961 - cel.  
agure@bilkent.edu.tr

## Research Paper

# A Stemness and EMT Based Gene Expression Signature Identifies Phenotypic Plasticity and is A Predictive but Not Prognostic Biomarker for Breast Cancer

Muhammad Waqas Akbar<sup>1</sup>, Murat Isbilen<sup>1,2</sup>, Nevin Belder<sup>1</sup>, Secil Demirkol Canli<sup>1,3</sup>, Baris Kucukkaraduman<sup>1</sup>, Can Turk<sup>1</sup>, Ozgur Sahin<sup>1</sup>, Ali Osmay Gure<sup>1</sup>✉

1. Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey.
2. DNAFect Genetics Consulting R&D and Biotechnology Inc., Kocaeli, Turkey.
3. Molecular Pathology Application and Research Center, Hacettepe University, Ankara, Turkey.

✉ Corresponding author: agure@bilkent.edu.tr; Bilkent University, Department of Molecular Biology and Genetics, SB-238; Cankaya, Ankara, Turkey 06800.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2019.03.05; Accepted: 2019.11.03; Published: 2020.01.01

## Abstract

**Aims:** Molecular heterogeneity of breast cancer results in variation in morphology, metastatic potential and response to therapy. We previously showed that breast cancer cell line sub-groups obtained by a clustering approach using highly variable genes overlapped almost completely with sub-groups generated by a drug cytotoxicity-profile based approach. Two distinct cell populations thus identified were CSC(cancer stem cell)-like and non-CSC-like. In this study we asked whether an mRNA based gene signature identifying these two cell types would explain variation in stemness, EMT, drug sensitivity, and prognosis *in silico* and *in vitro*. **Main methods:** *In silico* analyses were performed using publicly available cell line and patient tumor datasets. *In vitro* analyses of phenotypic plasticity and drug responsiveness were obtained using human breast cancer cell lines. **Key findings:** We find a novel gene list (CNCL) that can generate both categorical and continuous variables corresponding to the stemness/EMT (epithelial to mesenchymal transition) state of tumors. We are presenting a novel robust gene signature that unites previous observations related either to EMT or stemness in breast cancer. We show *in silico*, that this signature perfectly predicts behavior of tumor cells tested *in vitro*, and can reflect tumor plasticity. We thus demonstrate for the first time, that breast cancer subtypes are sensitive to either Lapatinib or Midostaurin. The same gene list is not capable of predicting prognosis in most cohorts, except for one that includes patients receiving neo-adjuvant taxane therapy. **Significance:** CNCL is a robust gene list that can identify both stemness and the EMT state of cell lines and tumors. It can be used to trace tumor cells during the course of phenotypic changes they undergo, that result in altered responses to therapeutic agents. The fact that such a list cannot be used to identify prognosis in most patient cohorts suggests that presence of factors other than stemness and EMT affect mortality.

Key words: Breast cancer, predictive biomarkers, tumor plasticity, transcriptomics.

## Introduction

Breast cancer (BC) is the second leading cause of mortality after lung cancer in women<sup>1</sup>. Currently the molecular classification of BC is based on expression of Estrogen Receptor (ER), Progesterone Receptor (PR) and Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2). Another classification, which is also based on gene expression differences identifies luminal A, luminal B, Her2 enriched, basal, and normal like subgroups, and provides a better prediction of prognosis and drug

response<sup>2,3</sup>. Some studies identified a cancer stem cell (CSC)-like subpopulation in BC, which was suggested as being responsible for metastasis and relapse of disease<sup>4,5</sup>. In general, CSCs are generally defined as CD44<sup>+</sup>/CD24<sup>-</sup> cells that possess the capacity of self-renewal<sup>6</sup>. The CSC hypothesis suggests - by definition - that non-CSC cells would be generated from CSC cells. On the other hand, a related but distinct mechanism, i.e. epithelial to mesenchymal

<http://www.jcancer.org>

transition (EMT), describes a reversible process through which epithelial cells transform into a mesenchymal state with the associated loss of epithelial features such as cell to cell contact and intracellular tight junctions, and the gain of mesenchymal features such as fibroblastoid morphology, increased motility and metastatic capability<sup>7</sup>. Cells which undergo EMT show similar features to CSCs such as tumorigenesis, lack of differentiation, mammosphere formation and resistance to anti-cancer therapies, suggesting that these two might be defining the same phenotype<sup>8</sup>. Mesenchymal cells can switch to an epithelial state (MET) under certain conditions<sup>9,10</sup>. The presence of CSCs has been conclusively shown *in vivo*<sup>11</sup> and *in vitro*<sup>12,13</sup>. It is known that CSC-like cells can be generated or enriched via mammosphere formation<sup>14</sup> or the ectopic expression of Twist or Snail<sup>15</sup>. However, both of these are also known to induce EMT. It is not yet clear if cells defined as CSCs in BC are identical to mesenchymal cells, and whether if via EMT, epithelial cell populations in BC can revert back to a CSC-like state.

An important distinction between CSC-like and non-CSC-like cells (or epithelial and mesenchymal cells) is their differential response to anti-cancer therapy. Commonly used drugs like Paclitaxel and Doxorubicin have been shown to affect primarily non-CSC cells, thereby enriching CSC-like populations in tumors<sup>16,17</sup>. Therefore, for successful tumor therapy, it seems drugs which specifically target CSC-like cells are also required. Ultimately, by the use of two drugs alternately or in parallel, both CSC- and non-CSC like cells could be eliminated.

In this study, using cells classified according to a novel gene-list (CNCL), we were able to identify candidate drugs for both CSC- and non-CSC cell types in breast cancer and validated our findings in several datasets *in silico*, as well as *in vitro*. However, we were not able to observe a consistent difference in clinical outcome in the majority of BC patient cohorts whose tumors had been thus classified.

## Methods

### *In silico* analyses

For *in silico* gene expression analysis, microarray datasets were downloaded from genomic data hosting websites, ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) and Gene expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). Each dataset was RMA normalized using BRB array tools<sup>18</sup>. FPKM values were used from GSE73526 which is a next generation RNA-sequencing dataset. Cluster tree 3.0 program<sup>19</sup> was used to hierarchically cluster data

and heatmaps were generated using Java Treeview<sup>20</sup>. For both genes and samples, Euclidean distances were calculated using complete linkage. Datasets generated using Affymetrix U133 A, B or U133 plus2 or Illumina platforms were used. Details of the datasets used are given in **Supplementary Table 1**. For *in silico* drug cytotoxicity analysis, IC50 or normalized activity area (AA) values were downloaded from CCLE<sup>21</sup> and CGP<sup>22</sup>. We did not select patients in any of the cohorts utilizing any filters as we intended to test if CNCL could identify prognosis, independent of clinical characteristics.

### Cell culture conditions

Breast cancer cell lines, MDA-MB-157, MDA-MB-231, MDA-MB-453, MDA-MB-468, MDA-MB-436, ZR751, JIMT1, BT474, BT20, CAL51 and MCF7 were cultured in DMEM media, while HCC202, HCC1954, HCC70, HCC1143, HCC38, T47D and HCC1937 cell lines were cultured in RPMI media. Respective media was supplemented with 10% FBS, 1% 200mM L-Glutamine (Lonza) and 1% 10K/10K Penicillin – Streptomycin (Lonza) of the total volume. All cell lines were incubated at 5% CO<sub>2</sub> and 37 °C in humidified incubator.

### Mammosphere culture (3D culture)

MDA-MB-157 cells were cultured in 75cm<sup>2</sup> low attachment flasks (Corning) to generate mammospheres (3D culture) in 3 separate experiments according to a previously published protocol<sup>23</sup>. Briefly, to generate mammosphere media, serum free DMEM media was enriched with 1X B27 (Invitrogen), 10 ng/ml EGF (Sigma Aldrich), FGF 20 ng/ml (Sigma Aldrich), 2 µg/ml heparin (Sigma Aldrich), L-Glutamine (Lonza) and Penicillin – Streptomycin (Lonza). To initiate mammosphere cultures, cells grown in monolayers were detached and resuspended in mammosphere media. Cells were counted and 2x10<sup>5</sup> cells were cultured in 75cm<sup>2</sup> low attachment flasks. After 3 days of culture, cells formed mammospheres. These mammospheres were separated from suspension using a 40 µm cell strainer (BD). Spheres were retrieved from the strainer by inverting it over petri dish and washing with PBS. After which cells were replated in low attachment flasks. This was carried out for 6 passages.

### Quantitative RT-PCR

Total RNA was isolated using Trizol (Ambion) and was treated with DNase (Ambion) according to manufacturer's protocols. RNA was reverse transcribed by Revert-Aid first strand cDNA synthesis kit (Thermo Fisher Scientific) according to the supplier's protocol using random hexamer primers. Quantitative profiling of selected genes by

qRT-PCR was performed in triplicates using Light Cycler 480 (Roche) with iTaq Universal SYBR Green Supermix (Bio-Rad). Primers were designed using Primer3 online tool (<http://primer3.ut.ee/>) and then validated with NCBI Primer-Blast Tool (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Primer sequences are shown in **Supplementary Table 2**. GAPDH was used as endogenous reference control. All expression data were calculated using the  $\Delta\Delta CT$  Method<sup>24</sup>.

### Cell viability analyses

Cell viability was analyzed in quadruplicates using CellTiter-Glo® Luminescent Cell Viability Assay Kit (Promega) and CyQUANT Cell Proliferation Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. For drug cytotoxicity in 2D culture, 3000 cells were plated in each well of 96 well plates and after 24 hours were treated with drugs (Lapatinib, Midostaurin) at 10 different doses (Lapatinib: 50  $\mu$ M-0.001  $\mu$ M, Midostaurin: 5 $\mu$ M-0.0001  $\mu$ M). Percent cell viability was calculated 72 hours after drug treatment. These values were then used to draw dose-inhibition curves using 6 models as (3 parameter, 3 parameter Top 100, 3 parameter Bottom 0, 4 parameter Top 100 and 4 parameter Bottom 0) using R software. The model with the lowest standard error was used to calculate IC50 and Activity area values. Same procedures were used to calculate cytotoxicity for mammosphere cultures except mammosphere generated cells were cultured in 96 well ultra-low attachment plates in mammosphere media.

### siRNA knockdown and cytotoxicity change

Knockdown of ZEB1 and SNAI2 genes was performed using commercially available siRNAs, targeting the respective transcripts (D-006564-04 (SiZEB1) and D-017386-02 (SNAI2), Dharmacon Lafayette, CO, USA) along with control siRNA from Qiagen (CAT# SI03650318, Qiagen, Hilden, Germany) using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) in serum free Opti-MEM at a concentration of 20 nM according to the manufacturer's protocol. In brief, cells were seeded at a density of  $2 \times 10^5$  cells/well in complete medium. When the cells reached 50-60 % confluency, medium was replaced with medium without antibiotics. The siRNA-Lipofectamine 2000™ mixture was added to the medium and cells were incubated at 37°C and 5% CO<sub>2</sub> for 48 hours. Cells were collected 72 h after transfection to analyze changes in transcript levels of related genes using QRT-PCR. For cytotoxicity experiments with Lapatinib and Midostaurin MDA-MB-157 cells (3000 cells/well) were seeded in 96-well plates and

twenty-four hours after seeding, cells were transfected with the siZEB1 and siSNAI2 at a final concentration of 20 nM, using Lipofectamine 2000™ (Invitrogen, CA) transfection reagent according to the manufacturer's protocol. After 24 hours of transfection, cells were treated with at 4 concentrations for Lapatinib (1, 5, 10 and 50  $\mu$ M) and Midostaurin (0.1, 0.5, 1 and 5  $\mu$ M). Cell viability was assessed 72 hours after treatment with CellTiter-Glo Luminescent Cell Viability Assay (Promega, Madison, WI, USA) as per the manufacturer's instructions. All treatment groups were set up in quadruplicate.

### Statistical analysis

Different treatment groups were compared using "t-test" and graphs were generated using GraphPad Prism v 6.01 (GraphPad Software Inc.). For survival analysis, Cox proportional hazard regression and Log-Rank tests were performed using the "Survival" package in R<sup>25,26</sup>. Kaplan Meier analysis was performed using IBM SPSS Statistics v.19.

## Results

### Identification of the CSC/non-CSC gene list (CNCL)

We previously showed<sup>27</sup> that breast cancer cell line sub-groups obtained by a clustering approach using highly variable gene expression differences<sup>28</sup> overlapped with that generated by a cytotoxicity based approach<sup>16</sup>. Thus, we could classify all breast cancer cell lines either as CSC-like or non-CSC-like. To generate a robust gene list that could be used to separate these two groups *in silico* and *in vitro*, we identified the most differentially expressed ( $p < 0.05$ , by t-test) 200 genes among the CSC-like or non-CSC-like cells using both the GSE36139 (CCLE)<sup>21</sup> and E-MTAB-783 (CGP)<sup>22</sup> datasets, which contain 56 and 39 breast cancer cell lines, respectively. One hundred and thirty three probesets, corresponding to 97 genes were common to both datasets, and 129 had highly correlated expression patterns (**Supplementary Figures 1A and 1B**). Of these, we selected 15 genes/probesets (8 up and 7 down-regulated in CSC-like cell lines), which had a minimum fold-change of 3 and a p value below 0.0002 and showed strong intergenic correlation (**Supplementary Tables 3A and 3B**); thus forming the CSC/non-CSC gene list (CNCL). A hierarchical clustering analysis using CNCL shows two clearly distinct cell clusters in CCLE and CGP datasets (**Figure 1A**). These two clusters were designated cancer-stem-cell-like and mesenchymal (CS/M) and non-cancer-stem-cell-like and epithelial (NS/E) as explained below.

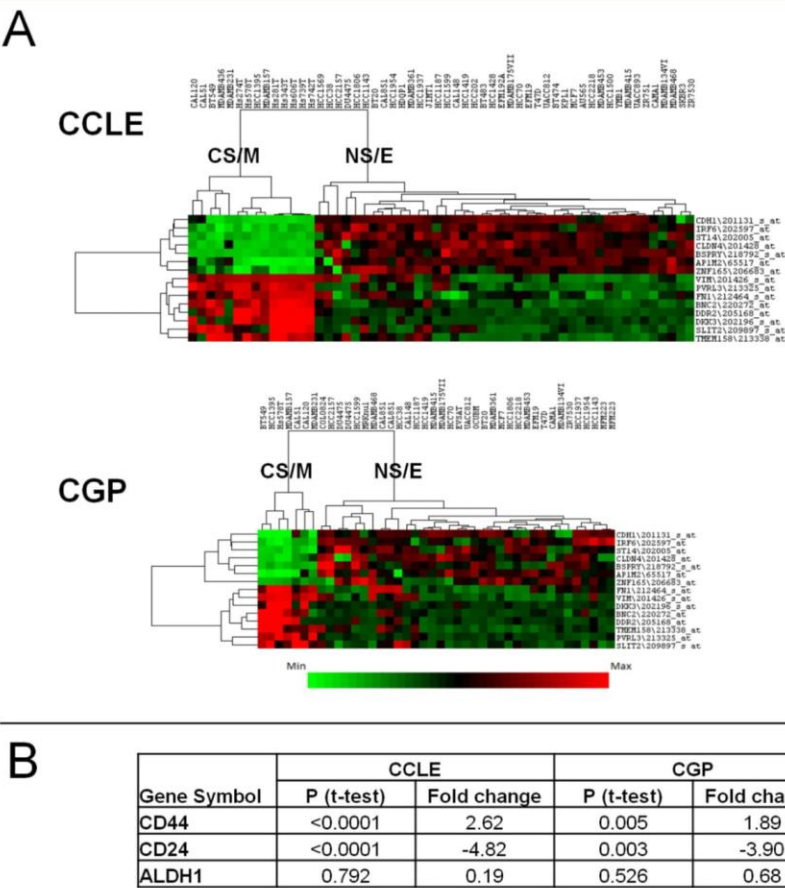


**Figure 1: Clustering of cell lines in CCLE and CGP datasets by CNCL.** CNCL generates two distinct clusters for CCLE and CGP cell lines (A). CS/M and NS/E classification correlates with stem cell markers CD44 and CD24 but not with ALDH1 in both datasets (see also **Supplementary Table 6**). CD44 and CD24 expression is significantly different between CSC-like and non-CSC-like clusters by t-test but ALDH1 did not show the same pattern in CCLE and CGP datasets (B).

We then used these cell clusters to perform gene-set-enrichment analyses (GSEA)<sup>29</sup>. Genesets enriched in both CSC-like and non-CSC-like groups were ranked by significance, and rank numbers obtained in both analyses were combined generating a ranksum value to identify commonly enriched genesets. As shown in **Supplementary Tables 4A and 4B**, gene sets indicative of a mesenchymal phenotype such as “Extracellular Matrix”, “Basement Membrane” and “Cell Migration” were enriched in CSC-like cells, while gene sets related to an epithelial phenotype like “Tight Junction” were enriched in

non-CSC-like cell lines, suggesting that the CSC-like phenotype is similar to a “mesenchymal”, and non-CSC-like to an “epithelial” phenotype. In support of this, an *in silico* analysis of CDH1 protein expression from two independent sources<sup>30,31</sup> showed that CDH1 was significantly up-regulated among non-CSC-like cell lines compared to those with a CSC-like phenotype (**Supplementary Figure 2**). We thus conclude that CNCL classifies cell lines into relatively epithelial and mesenchymal phenotypes and therefore name these two types of cells CS/M and NS/E.

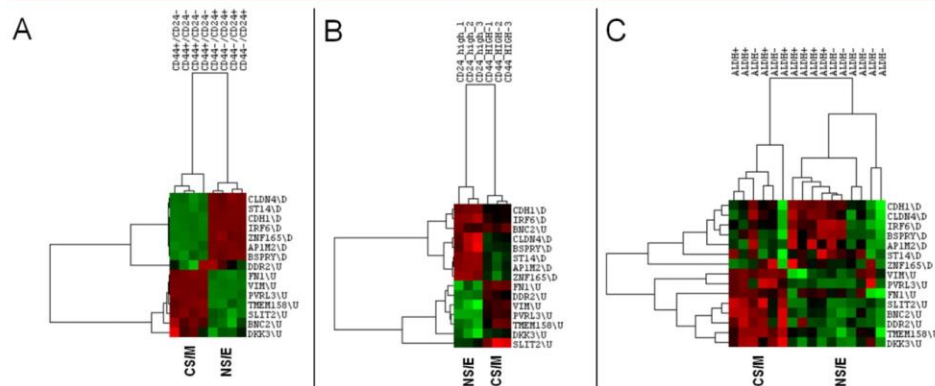


**Figure 2: CNCL identifies stemness in silico.** CNCL generates two clusters of CD44+/CD24- and CD44-/CD24+ MCF10A breast cancer cells (GSE15192) (A), as well as in Ras-transformed human mammary epithelial cells sorted based on their CD44 or CD24 expression (GSE36643) (B). In primary breast cancer cells sorted for ALDH+ expression (GSE52327), CNCL based differentiation is not clear (C).

To test the robustness of the CNCL gene list we asked if cell lines defined as CS/M or NS/E in CGP and CCLE would be classified consistently by CNCL in other datasets as well. Cell line gene expression data from 3 datasets: GSE24717<sup>32</sup> and GSE50811<sup>33</sup> (microarray) and GSE73526<sup>31</sup> (RNA sequencing) was used for this purpose. Of 123 cell line assignments, only 1 (HCC1419 in GSE50811) was classified inconsistently across datasets. We therefore, conclude that CNCL behaves consistently across different platforms and is unaffected from inter-laboratory variation (Supplementary Table 5).

We then asked if the CNCL gene list could provide biologically significant distinctions among cell and tissue subtypes *in silico* and *in vitro*.

#### CNCL reflects stemness in silico

NS/E and CS/M cells identified by CNCL showed significantly different levels of expression, *in silico* for CD24 and CD44 in both CCLE and CGP datasets, while another stem-cell marker, ALDH did not show differential expression among cell groups (Figure 1B, and Supplementary Table 6). CNCL could clearly identify CD44+/CD24- stem cells and CD44-/CD24+ non stem cells derived from MCF10A breast cancer cells (GSE15192)<sup>34</sup> (Figure 2A). Similarly, CNCL could distinguish stem and non-stem HMLER cells (Ras-transformed human mammary epithelial cells) obtained by CD44/CD24 expression sorting (GSE36643)<sup>35</sup> (Figure 2B). The distinction between CS/M and NS/E type cells in primary breast cancer cells (GSE52327)<sup>36</sup> was less clear (Figure 2C) possibly reflecting the heterogeneity of these cells. ALDH expression was not related to CNCL defined phenotypes in this analysis as well.

Expression of all 15 genes in CNCL correlated with CD24 and CD44 but not with ALDH in GSE15192 (Supplementary Table 7A) and in GSE36643 (Supplementary Table 7B), similar to CCLE or CGP. Cumulatively, these data suggest NS/E cells are highly similar non-stem cancer cells, while the opposite is true for CS/M cells.

#### CNCL can identify EMT in silico

We next analyzed the relation of CNCL determined clusters with those determined in the context of EMT. When the GSE9691 dataset that contains expression data from HMLER cells in which E-Cad gene was downregulated either by shRNA treatment or by the transfection of a dominant negative version of E-Cad, or cells in which both E-Cad and  $\beta$ -Cat were downregulated<sup>37</sup> were analyzed, CNCL correctly classified all E-Cad down-regulated cells as CS/M, and all control cells, and those where effects of E-Cad downregulation are reversed by simultaneous  $\beta$ -Cat downregulation, as NS/E cells (Figure 3A). In a separate analysis, HMLER cells that were retrovirally transduced with TGF $\beta$ , Twist, Gsc or Snail inducing EMT<sup>38</sup>, were identified by CNCL as CS/M cells (GSE24202) (Figure 3B). Unexpectedly when E-cad expression was downregulated by shRNA in these cells, they were still identified as NS/E. However, shRNA dependent downregulation of E-cad in these cells was significantly less efficient compared to the other experiments (TGF $\beta$ , Twist, Snail and Gsc), suggesting that a conversion to a more mesenchymal state possibly requires E-cad downregulation below a certain threshold (Figure 3B, bottom). To test if CNCL could classify primary tumor cells we analyzed the

GSE7515 dataset<sup>39</sup>. Of 11 primary tumor cells, 10 were classified either as NS/E or intermediary (I) phenotypes by CNCL, while 1 was of a CS/M phenotype. In contrast, among 15 mammospheres, generated from primary tumor cells, in which EMT would be expected to occur, 12 were characterized as CS/M, and 3 as either NS/E or I cells (Figure 3C). Therefore, 22 of 26 primary cell and mammosphere cultures were clustered by the CNCL list in an expected fashion. Exceptions could be due to the fact that primary cells can have a mesenchymal phenotype to start with, and because *in vitro* mammosphere cultures might not always generate mesenchymal cells. Cumulatively these results indicate that CNCL is able to distinguish cells with relatively epithelial features from those with mesenchymal features for both cell lines and primary tumor cells.

#### CNCL can predict drug sensitivity *in silico*

As both cancer stem cells and mesenchymal cancer cells have been reported to resist the majority of cancer chemotherapy and/or targeted therapy regimens, we asked if CNCL could identify drug sensitive and resistant cells *in silico*. As shown in Figure 4A, CNCL could clearly distinguish Doxorubicin sensitive MCF7 cells from those that were made resistant to this drug via prolonged exposure (GSE24460)<sup>17</sup>. Development of drug resistance *in vivo* is a major cause of therapy failure. To investigate if the CNCL list could identify a phenotype switch in *in vivo* drug resistance, we utilized the GSE10281 dataset which contains pre- and post- Letrozol treated tumor tissue from 18 patients<sup>39</sup>. CNCL based analysis of this dataset revealed that after a 3 month treatment period, out of 14 tumors with a pre-treatment NS/E phenotype, 7 had converted to CS/M tumors post-therapy. And among 4 tumors with a CS/M phenotype, pre-therapy, one converted to a NS/E tumor post-therapy (Figure 4B). We interpret these results to mean that most tumors with a NS/E phenotype pre-therapy, are prone to generate CS/M tumors post-therapy, and thus possibly become drug resistant.

#### Analyses based on a quantitative stemness score (SS)

We generated two matrices, one for CS/M related-, and one for NS/E related genes (Supplementary Table 8) based on the median expression levels for each gene in the CNCL for CS/M and also for NS/E cells in the CCLE database after mean standardization. We then correlated each tumor sample's standardized CNCL gene expression values to both the CS/M and NS/E matrices generating two correlation values (Pearson's *r*) for each matrix,

CS/M(*r*) and NS/E(*r*), respectively. Analysis of cells based on their correlation scores clearly distinguished Paclitaxel sensitive and resistant MDA-MB-231 cells in GSE12791, as well as cells re-sensitized to Paclitaxel through differentiation inducing Bexarotene treatment<sup>40</sup> (Figure 5A). In another case (GSE23399), we found that the same analysis was able to reveal the phenotypic change observed in cancer-associated fibroblasts (CAF) in response to paclitaxel therapy<sup>41</sup>. As shown in Figures 5B, the longer CAFs were exposed to Paclitaxel, the more CS/M-like, and less NS/E-like they became. A similar switch to a more CS/M-like phenotype was also observed in BT20 breast cancer cells made resistant to a targeted therapy agent, Lapatinib (GSE16179) (Figure 5C)<sup>42</sup>. Next we generated the SS (Delta (*r*)) by calculating the difference of two Pearson *r* values for each sample (i.e.: CS/M (*r*) - NS/E (*r*)), where negative values indicate a more NS/E phenotype. Using SSs we found that tumor tissues from 29 breast cancer patients obtained before Anthracycline and Taxane based neo-adjuvant chemotherapy were more NS/E-like, compared to post-therapy tumor tissues from 32 patients (GSE28844)<sup>43</sup>, suggesting that chemotherapy exposure had resulted in a general phenotype switch in most of these tumors (Figure 5D). Therefore, just like the hierarchical clustering based experiments summarized in Figure 4, the SS based approach could identify cells and tumor tissues which responded to chemotherapy treatment. Based on these results, we conclude that CNCL can trace tumor phenotype plasticity.

#### Utilization and *in vitro* validation of CNCL as a drug sensitivity predictor

To identify drugs which could preferentially target CS/M and NS/E cells, we analyzed the CCLE database and found 4 drugs that affect these subtypes differentially: Midostaurin and Elesclomol were significantly more effective on CS/M cells, whereas Lapatinib and Panabinoestat were cytotoxic for NS/E cells (Figure 5E). To validate these *in silico* findings *in vitro*, we first tested whether the CNCL based classifications could be reproduced using qPCR generated data. Indeed, qPCR data obtained for genes in the list were highly concordant with CCLE (GSE36139) microarray expression data (Supplementary Table 9A), and inter-gene correlations were as expected (Supplementary Table 9B). We then generated a SS matrix based on qPCR data, similar to that generated for *in silico* analyses: BC cell lines were hierarchically clustered using standardized qPCR expression data thus generating CS/M and NS/E cell groups. Each group was used

independently to calculate average expression values for all 15 genes which were then used to generate the CS/M and NS/E matrices and ultimately SSs (delta (r) values).

*In vitro*, Midostaurin was preferentially cytotoxic for CS/M cells in line with our *in silico* findings. We

observed that Lapatinib was more effective on NS/E cells (Figure 6A and Supplementary Figure 3). We were unable to show a differential effect for Elesclomol and Panobinostat on CS/M vs. NS/E cells *in vitro* (data not shown).

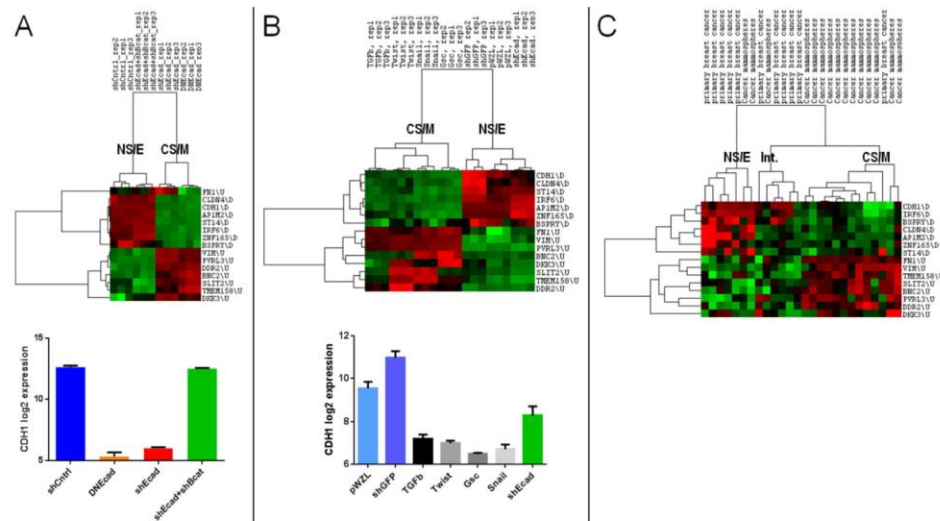


**Figure 3: CNCL identifies EMT *in silico*.** CNCL identifies cells in which CDH1 was downregulated using either shRNA (shEcad) or via dominant-negative CDH1 expression (DNEcad), as CS/M cells, when compared to controls (shCntrl) or cells where both CDH1 and CTNNB1 were downregulated (shEcad+shBcat) which are of the NS/E phenotype (GSE9691). CDH1 expression was significantly lower in CS/M as compared to NS/E cells (CDH1 vs. shCntrl: <0.0001, CDH1 vs. shCDH1+shBcat (CTNNB1) <0.0001) (A). In GSE24202, cells over-expressing TGF- $\beta$ , Twist, Gsc or Snail are of the CS/M phenotype compared to controls (shGFP and pWZL). Although cells treated with shCDH1 downregulated CDH1, its expression remained higher than TGF- $\beta$  (p=0.02), Twist (p<0.01), Gsc (p<0.01), and Snail (p<0.01) (B). For primary breast cancer cells and mammospheres generated from those samples CNCL identified most of the mammospheres (12/15) as CS/M and all but one primary cancer samples as NS/E and intermediary, in GSE7515 (C).

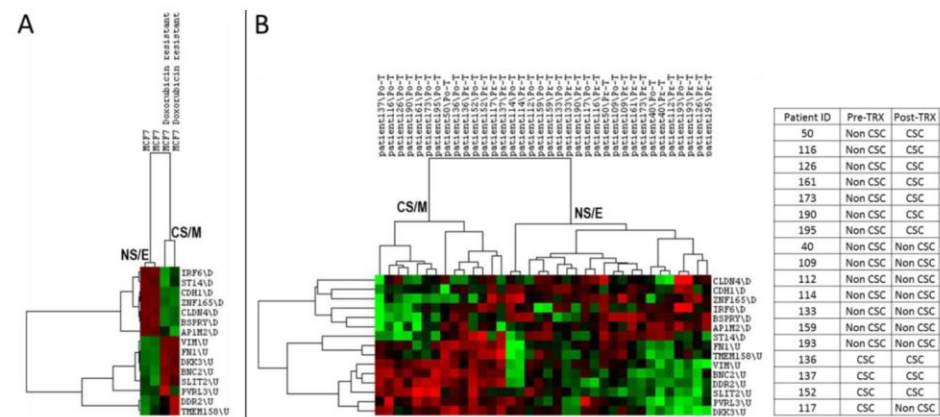


**Figure 4: CNCL can predict drug sensitivity *in silico*.** CNCL identifies doxorubicin resistant MCF7 cells as CS/M, compared to control cells (GSE24460) which are NS/E (A). (B): CNCL identifies Letrozol resistance in pre- and post- Letrozol treated tumor tissue (GSE10281) from 14 NS/E tumors, 7 converted to a CS/M phenotype, post treatment.

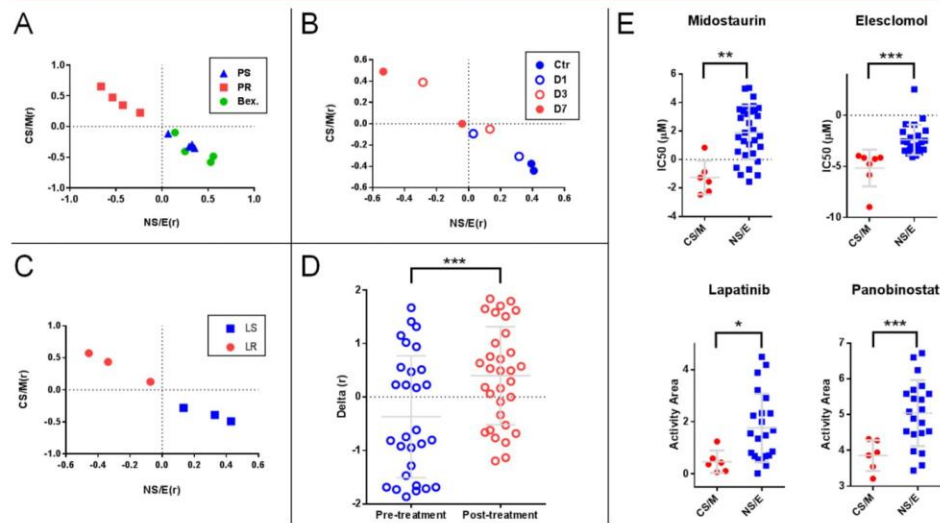


**Figure 5: Stemness score (SS) based analyses show that the CNCL can reflect tumor plasticity.** **A:** (GSE12791) Paclitaxel resistant MDA-MB-231 cells (PR) show high CS/M (r) and low NS/E (r) as compared to Paclitaxel sensitive cells (PS) or those PR cells treated with Bexarotene (Bex). **B:** (GSE23399) Paclitaxel treated carcinoma associated fibroblasts (CAFs) treated for 3 and 7 days (D3 and D7) show higher CS/M (r) and low NS/E (r) when compared to cells treated for 1 day or control cells (D1 and Ctr). **C:** (GSE16179) Lapatinib-resistant BT474 cell line (LR) show high CS/M (r) and low NS/E (r) when compared to sensitive cells (LS). **D:** (GSE28844) Patient tumors post-anthracycline and taxane chemotherapy, show high Delta (r) when compared to pre-treatment where Delta (r): [CS/M(r)] - [NS/E(r)]. **E:** Midostaurin and Elesclomol are cytotoxic preferentially for CS/M while Lapatinib and Panobinostat target NS/E. \* $p < 0.05$  (t-test), \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

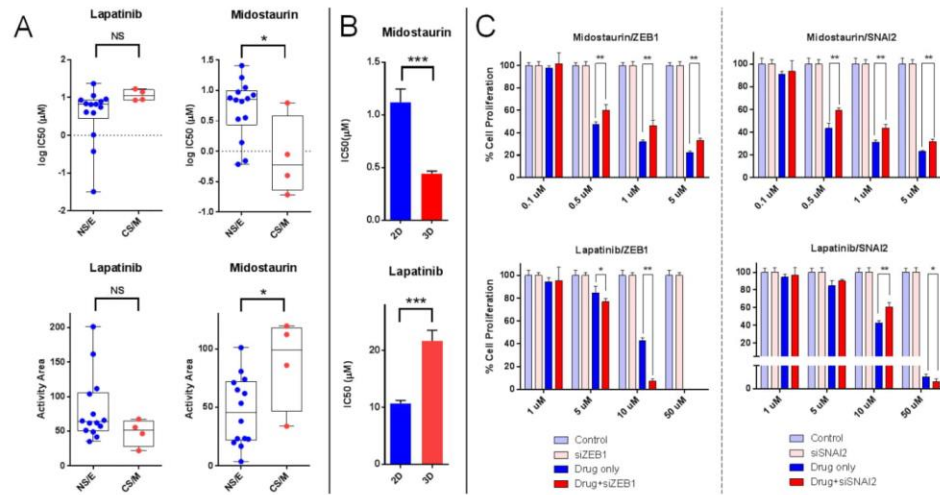


**Figure 6: In vitro validation of CNCL as a drug sensitivity predictor.** **A:** Midostaurin is preferentially cytotoxic for CS/M cells compared to NS/E cells, while the reverse is observed for Lapatinib (CyQUANT assay). **B:** MDA-MB-157 spheroids showed increased sensitivity to Midostaurin and resistance to Lapatinib. **C:** ZEB1 and SNAI2 knock down results in decreased sensitivity to Midostaurin and increased sensitivity to Lapatinib in MDA-MB-157 cells. \* $p < 0.05$  (t-test), \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , NS: not significant.

To test if the CNCL signature reflected tumor phenotype plasticity *in vitro* and whether this would relate to a change in drug sensitivity, we generated spheroids from MDA-MD-157 cells and measured gene expression by qPCR, and the simultaneous change in drug sensitivity. The qPCR based initial SS for MDA-MD-157 cells (-0.145) increased to +0.145 upon spheroid formation, reflecting a shift towards a CS/M phenotype. This happened concurrently with a significant increase in sensitivity to Midostaurin and a decrease in sensitivity to Lapatinib (Figure 6B). Therefore, spheroid generation from MDA-MD-157 cells caused a change in phenotype that could be quantitated by the CNCL, concurrent with a change in drug sensitivity *in vitro*.

As ZEB1 and SNAI2 are considered critical genes in maintaining a mesenchymal/CSC-like phenotype, we knocked down both genes individually in MDA-MD-157 (Supplementary Figure 4) and determined changes in CNCL gene expression and drug sensitivity. SS Delta (r) values changed from +5.1 in control siRNA transfected cells, to -0.12 and -0.01 in ZEB1 and SNAI2 siRNA transfected cells, respectively, indicating a switch from a CS/M to a NS/E phenotype. In parallel to this change in phenotype, we observed significantly increased sensitivity to Lapatinib and decreased sensitivity to Midostaurin (Figure 6C). The fact that the change in sensitivity to Lapatinib was more obvious for ZEB1 knock-down cells is likely due to the fact that these underwent a stronger shift towards a NS/E phenotype, compared to SNAI2 knock-down cells.

Cumulatively, these results indicate that tumor cells that are innately of a CS/M phenotype, or those that are induced to become so, are less sensitive to Lapatinib and more sensitive to Midostaurin, compared to cells that are innately NS/E or are induced to become so, which show the opposite pattern of sensitivity.

#### Evaluation of CNCL as a prognostic tool for breast cancer

To determine if a CNCL based classification of tumor samples could be used to predict prognosis, we generated SS values for all samples within several breast cancer datasets: Metabric cohorts (British (n=994) and Canadian (n=997))<sup>44</sup>, GSE1456 (n=159)<sup>45</sup>, GSE2034 (n=286)<sup>46</sup>, GSE2603 (n=82)<sup>47</sup>, GSE3494 (n=251)<sup>48</sup>, GSE4922(n=249)<sup>49</sup>, GSE6532 (n=380)<sup>50</sup>, GSE7390 (n=198)<sup>51</sup>, GSE11121(n=200)<sup>52</sup>, GSE12276 (n=204)<sup>53</sup>, GSE19615 (n=115)<sup>54</sup>, GSE20685(n=327)<sup>55</sup>, GSE21653 (n=266)<sup>56</sup>, GSE25066(n=508)<sup>57</sup> and GSE58812 (n=107)<sup>58</sup>. We included all cohorts with survival data and did not use filtering criteria as our

aim was to discover any prognostic association the CNCL had, independent of confounding factors. As shown in Supplementary Table 10, SS delta (r) based evaluation of prognosis of 15 patient cohorts in 29 tests with various end-point measures revealed statistically significant differences between CS/M and NS/E groups in only 7 tests. And in only two of these (GSE25066 and GSE2603) patients with primarily CS/M tumors had worse prognosis compared to NS/E tumor-harboring patients. Interestingly the GSE25066 cohort, where this trend was most obvious, consisted of patients who had received Taxane (incl. Paclitaxel) based neo-adjuvant therapy. Patients in this cohort with NS/E tumors had longer distant relapse free survival compared to those with CS/M tumors (Figure 7A, B). Although SS distributions were significantly different for PAM50 subtypes in this cohort (Figure 7C), the CNCL based cut-off was a statistically significant prognostic factor in a multivariate analysis that included N stage, T stage and PAM50 subtypes (Figure 7D). In line with our results summarized in Figure 5, these show a clear relation between Paclitaxel sensitivity and CNCL identified groups, results obtained for this cohort suggest that the CNCL list might be especially beneficial in determining outcome in patients receiving Taxane based treatment as neo-adjuvant therapy. In 22 tests, we did not observe a significant association between SSs and survival end-measures, and in 5 tests we observed the opposite of what we expected, a direct relation between better prognosis and the CS/M phenotype. Given the very consistent results we obtained for the experiments summarized in the previous sections, we think that the lack of a clear relation between CS/M and worse prognosis in patient cohorts is due to the heterogeneity of tumors, and also because information related to chemotherapy or other treatment was not available for most samples.

#### Discussion

There exists a large array of markers that have been reported to identify different subtypes of breast cancer cells<sup>59</sup>. We find that CNCL correlates with CD24 and CD44 expression but not with ALDH. EMT markers for breast cancer seem to be shared to a greater extent among cancers and can be used to classify tumors across tissue-types<sup>60</sup>. The CNCL gene list defined in this study is related to a previously defined EMT gene list<sup>60</sup> as indicated by the high concordance between the two (data not shown). Similar to the EMT gene list, CNCL can classify both primary tumor tissues as well as cell lines.

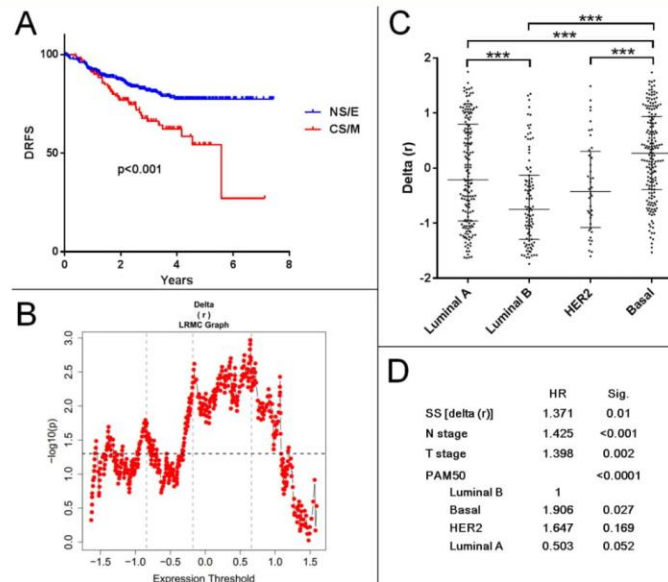


**Figure 7: CNCL can predict prognosis for patients treated with Taxane based neo-adjuvant therapy.** A: (GSE25066) Patients with CS/M tumors show worse prognosis compared to NS/E group, when a SS cut-off of 0.6459 is used. B: Log rank test with multiple cut-offs (LRMC). C: Delta (r) value distribution among breast cancer subtypes. D: Multivariate analysis with multiple clinical parameters showing SS is an independent prognostic factor for patients with paclitaxel neo-adjuvant treatment. \*\*\*p<0.001.

The differentiation state of a tumor is accepted as the key determinant of resistance to therapy<sup>61</sup>. However, tumors have the capability of altering their phenotype, which is referred to as phenotype plasticity. This term is used to define both de-differentiation as well as trans-differentiation in cancer cells<sup>62</sup>. Our findings show that phenotype plasticity occurs almost without exception in tumor cells, *in vivo* and *in vitro*, in response to chemotherapy and CNCL is able to dynamically follow the phenotype switch between these two states. As both Zeb1 and Snail1 induction have been associated with EMT and drug resistance<sup>63</sup>, and as Zeb1 is also a critical component of tumor cell plasticity<sup>64</sup>, it is not surprising that the down-regulation of these would correlate with increased sensitivity to anti-cancer agents. Our results clearly show that the CNCL can identify tumor states and cell lines undergone through EMT or stemness as a result of phenotype plasticity. CNCL is therefore, a novel gene list that can account for both stemness and EMT in breast cancer, indicating that the two are related when analyzed in the context of this signature.

Metastasis is considered to be the major reason for tumor related mortality. Although EMT clearly

has a role in tumor progression and metastasis, recent data suggests that the metastatic potential of tumor cells and their ability to undergo EMT are not necessarily related<sup>65</sup>. These studies suggest that instead of classifying tumors either as epithelial or mesenchymal, to determine the rate at which plasticity occurs might be a better indicator of prognosis and drug responses<sup>62</sup>. It is therefore, important to be able to quantify the exact phenotypic status of a tumor within the EMT scale and to be able to trace its change. In this line, we believe the CNCL defined here is worth validating in larger studies.

Gene signatures in BC have been shown to be able to predict sensitivity to various drugs *in vitro*<sup>66</sup> and EMT has been reported to result in resistance to the EGFR inhibitor Gefitinib<sup>67</sup>. Our results suggest the best strategy for BC treatment is possibly the utilization of two drugs simultaneously, each which would target either NS/E or CS/M cells. In line with our earlier findings<sup>27</sup>, we show here for the first time *in vitro*, that Lapatinib and Midostaurin are candidates for such combination therapy, as they target selectively either NS/E or CS/M cells, respectively. Even though EMT might be a signature shared among tumor types, combination treatment

suggestions obtained for breast cancer in this study are different from those we found for other tumor types, such as colon cancer and melanoma<sup>68,69</sup>. One explanation for this can be that NS/E and CS/M cell types defined for tumors originating from different tissues actually define different phenotypes in the EMT or plasticity scale. Indeed most CS/M cells of colon tumors are similar to NS/E cells of breast cancer, while melanoma cells are among the most CS/M cells when compared to all other tumors<sup>60</sup>.

Our results clearly show that most tumors with a pre-therapy NS/E phenotype will switch to a CS/M phenotype post-therapy. This might explain why most tumors are resistant to secondary treatment. Changes in gene expression following NACT have shown that highly proliferative tumors are more sensitive to chemotherapy<sup>70,71</sup>. Here we show that the CNCL gene list can also robustly distinguish patients who will have a better prognosis post taxane-based NACT. However, in light of our findings summarized here and those of others<sup>72</sup>, it should be noted that NACT might induce a CS/M phenotype in tumors, reducing drug responsiveness and thus, overall or disease-free survival. Therefore, it is critical to conclusively determine if NACT is actually decreasing survival of patients whose tumors switch to a CS/M state, following therapy, rather than improving survival. The CNCL could be utilized to follow tumor plasticity in response to such therapies to evaluate responses to NACT and to help answer such questions.

## Conclusions

We show in this study for the first time that the CNCL gene list, which includes stem-cell and EMT features, reflects tumor plasticity and thus a shift in cytotoxicity profiles, especially in response to Midostaurin and Lapatinib in breast cancer; and that it is applicable both *in silico* and *in vitro*. Although the gene list is not uniformly predictive of prognosis, it can be utilized to identify patients who might benefit from Taxane-based neo-adjuvant therapy.

## Abbreviations

CSC: Cancer stem cells; BC: Breast cancer; PR: Progesterone receptor; ER: Estrogen receptor; EMT: Epithelial to mesenchymal transition; MET: Mesenchymal to epithelial transition; CS/M: Cancer-stem-cell-like and mesenchymal; NS/E: Non cancer stem cell like and epithelial; GSEA: Gene set enrichment analysis; CNCL: CSC/non-CSC gene list.

## Supplementary Material

Supplementary figures and tables.  
<http://www.jcancer.org/v11p0949s1.pdf>

## Acknowledgements

This work was partially funded by a grant from the Turkish Scientific and Technological Research Council/TUBITAK (117S058) to AOG and one from The Higher Education Commission of Pakistan to MWA.

## Contributions

**MWA:** conceptualization, methodology, validation, data analysis, writing original manuscript. **MI:** conceptualization, data analysis. **NB:** acquisition of data. **SD:** formal analysis. **BK:** acquisition of data. **CT:** acquisition of data. **OS:** conceptualization and design, development of methodology, writing review. **AOG:** conceptualization, methodology, data analysis, writing original manuscript, acquisition of funds, study supervision.

## Competing Interests

The authors have declared that no competing interest exists.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7-30.
2. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-536.
3. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160-1167.
4. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 2003;100:3983-3988.
5. Balic M, Lin H, Young L, et al. Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype. *Clin Cancer Res* 2006;12:5615-5621.
6. Clarke MF, Dick JE, Dirks PB, et al. Cancer stem cells—perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res* 2006;66:9339-9344.
7. Wang Y, Zhou BP. Epithelial-mesenchymal transition in breast cancer progression and metastasis. *Chin J Cancer* 2011;30:603-611.
8. Floor S, van Staveren WC, Larsson D, et al. Cancer cells in epithelial-to-mesenchymal transition and tumor-propagating-cancer stem cells: distinct, overlapping or same populations. *Oncogene* 2011;30:4609-4621.
9. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* 2009;9:265-273.
10. Yilmaz-Ozcan S, Sade A, Kucukkaraduman B, et al. Epigenetic mechanisms underlying the dynamic expression of cancer-testis genes, PAGE2, -2B and SPANX-B, during mesenchymal-to-epithelial transition. *PLoS One* 2014;9:e107905.
11. Gilbertson RJ, Graham TA. Cancer: Resolving the stem-cell debate. *Nature* 2012;488:462-463.
12. Fillmore CM, 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.
13. Charafe-Jauffret E, Ginestier C, Iovino F, et al. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer Res* 2009;69:1302-1313.
14. Grimshaw MJ, Cooper L, Papazisis K, et al. Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. *Breast Cancer Res* 2008;10:R52.
15. Mani SA, Guo W, Liao MJ, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008;133:704-715.
16. Gupta PB, Onder TT, Jiang G, et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 2009;138:645-659.

17. Calcagno AM, Salcido CD, Gillet JP, et al. Prolonged drug selection of breast cancer cells and enrichment of cancer stem cell characteristics. *J Natl Cancer Inst* 2010;102:1637-1652.
18. Simon R, Lam A, Li MC, et al. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform* 2007;3:11-17.
19. de Hoon MJ, Imoto S, Nolan J, et al. Open source clustering software. *Bioinformatics* 2004;20:1453-1454.
20. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics* 2004;20:3246-3248.
21. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483:603-607.
22. Garnett MJ, Edelman EJ, Heidorn SJ, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;483:570-575.
23. Balic M, Schwarzenbacher D, Stanzer S, et al. Genetic and epigenetic analysis of putative breast cancer stem cell models. *BMC Cancer* 2013;13:358.
24. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402-408.
25. Team RC. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria ISBN 2007.
26. Therneau TM, Grambsch PM. *Modeling survival data: extending the Cox model*. Springer Science & Business Media, 2013.
27. Isbilen M, Senses KM, Gure AO. Predicting Chemotherapy Sensitivity Profiles for Breast Cancer Cell Lines with and Without Stem Cell-Like Features. *Current Signal Transduction Therapy* 2013;8:268-273.
28. Kao J, Salari K, Bocanegra M, 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.
29. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-15550.
30. Prat A, Adamo B, Cheang MC, et al. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist* 2013;18:123-133.
31. Marcotte R, Sayad A, Brown KR, et al. Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* 2016;164:293-309.
32. Shats I, Gatz ML, Chang JT, et al. Using a stem cell-based signature to guide therapeutic selection in cancer. *Cancer Res* 2011;71:1772-1780.
33. Dezo Z, Oestreicher J, Weaver A, et al. Gene expression profiling reveals epithelial mesenchymal transition (EMT) genes can selectively differentiate eribulin sensitive breast cancer cells. *PLoS One* 2014;9:e106131.
34. Bhat-Nakshatri P, Appaiah H, Ballas C, et al. SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype. *BMC Cancer* 2010;10:411.
35. Battula VL, Shi Y, Evans KW, et al. Ganglioside GD2 identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest* 2012;122:2066-2078.
36. Liu S, Cong Y, Wang D, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* 2014;2:78-91.
37. Onder TT, Gupta PB, Mani SA, et al. Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res* 2008;68:3645-3654.
38. Taube JH, Herschkowitz JL, Komurov K, 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 U S A* 2010;107:15449-15454.
39. Creighton CJ, Li X, Landis M, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A* 2009;106:13820-13825.
40. Luo W, Schork NJ, Marschke KB, et al. Identification of polymorphisms associated with hypertriglyceridemia and prolonged survival induced by bexarotene in treating non-small cell lung cancer. *Anticancer Res* 2011;31:2303-2311.
41. Chan TS, Hsu CC. Metronomic chemotherapy prevents therapy-induced stromal activation and induction of tumor-initiating cells. *J Exp Med* 2016;213:2967-2988.
42. Liu L, Greger J, Shi H, et al. Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. *Cancer Res* 2009;69:6871-6878.
43. Vera-Ramirez L, Sanchez-Rovira P, Ramirez-Tortosa CL, et al. Transcriptional shift identifies a set of genes driving breast cancer chemoresistance. *PLoS One* 2013;8:e53983.
44. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346-352.
45. Pawitan Y, Bjohle J, Amler L, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005;7:R953-964.
46. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671-679.
47. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. *Nature* 2005;436:518-524.
48. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 2005;102:13550-13555.
49. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 2006;66:10292-10301.
50. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25:1239-1246.
51. Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 2007;13:3207-3214.
52. Schmidt M, Bohm D, von Tonne C, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 2008;68:5405-5413.
53. Bos PD, Zhang XH, Nadal C, et al. Genes that mediate breast cancer metastasis to the brain. *Nature* 2009;459:1005-1009.
54. Li Y, Zou L, Li Q, et al. Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat Med* 2010;16:214-218.
55. Kao KJ, Chang KM, Hsu HC, et al. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* 2011;11:143.
56. Sabatier R, Finetti P, Adelaide J, et al. Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One* 2011;6:e27656.
57. Hatzis C, Pusztai L, Valero V, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *Jama* 2011;305:1873-1881.
58. Jezequel P, Loussouarn D, Guerin-Charbonnel C, et al. Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res* 2015;17:43.
59. Bai X, Ni J, Beretov J, et al. Cancer stem cell in breast cancer therapeutic resistance. *Cancer Treat Rev* 2018;69:152-163.
60. Tan TZ, Miow QH, Miki Y, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med* 2014;6:1279-1293.
61. Housman G, Byler S, Heerboth S, et al. Drug resistance in cancer: an overview. *Cancers (Basel)* 2014;6:1769-1792.
62. Gupta PB, Pastushenko I, Skibinski A, et al. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* 2019;24:65-78.
63. Shang Y, Cai X, Fan D. Roles of epithelial-mesenchymal transition in cancer drug resistance. *Curr Cancer Drug Targets* 2013;13:915-929.
64. Krebs AM, Mitschke J, Lasierra Losada M, et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat Cell Biol* 2017;19:518-529.
65. Pastushenko I, Brisebarre A, Sifrim A, et al. Identification of the tumour transition states occurring during EMT. *Nature* 2018;556:463-468.
66. Hsu YC, Chen HY, Yuan S, et al. Genome-wide analysis of three-way interplay among gene expression, cancer cell invasion and anti-cancer compound sensitivity. *BMC Med* 2013;11:106.
67. Weng CH, Chen LY, Lin YC, et al. Epithelial-mesenchymal transition (EMT) beyond EGFR mutations per se is a common mechanism for acquired resistance to EGFR TKI. *Oncogene* 2019;38:455-468.
68. Demirkol S, Gomceli I, Isbilen M, et al. A Combined ULBP2 and SEMA5A Expression Signature as a Prognostic and Predictive Biomarker for Colon Cancer. *J Cancer* 2017;8:1113-1122.
69. Senses KM, Ghasemi M, Akbar MW, et al. Phenotype-based variation as a biomarker of sensitivity to molecularly targeted therapy in melanoma. *Medchemcomm* 2017;8:88-95.
70. Klintonman M, Buus R, Cheang MC, et al. Changes in Expression of Genes Representing Key Biologic Processes after Neoadjuvant Chemotherapy in Breast Cancer, and Prognostic Implications in Residual Disease. *J Exp Med* 2016;222:2405-2416.

71. Klintman M, Buus R, Cheang MC, et al. Changes in Expression of Genes Representing Key Biologic Processes after Neoadjuvant Chemotherapy in Breast Cancer, and Prognostic Implications in Residual Disease. *Clin Cancer Res* 2016;22:2405-2416.
72. Liu L, Yang L, Yan W, et al. Chemotherapy Induces Breast Cancer Stemness in Association with Dysregulated Monocytosis. *Clin Cancer Res* 2018;24:2370-2382.



Cite this: *Med. Chem. Commun.*,  
2017, 8, 88

## Phenotype-based variation as a biomarker of sensitivity to molecularly targeted therapy in melanoma†

Kerem M. Senses,<sup>a</sup> Mehdi Ghasemi,<sup>a</sup> Muhammad W. Akbar,<sup>a</sup> Murat Isbilen,<sup>a</sup>  
Anna L. Fallacara,<sup>b</sup> Shoshana Frankenburg,<sup>c</sup> Silvia Schenone,<sup>d</sup> Michal Lotem,<sup>c</sup>  
Maurizio Botta<sup>b</sup> and Ali O. Gure<sup>\*a</sup>

Transcriptomic phenotypes defined for melanoma have been reported to correlate with sensitivity to various drugs. In this study, we aimed to define a minimal signature that could be used to distinguish melanoma sub-types *in vitro*, and to determine suitable drugs by which these sub-types can be targeted. By using primary melanoma cell lines, as well as commercially available melanoma cell lines, we find that the evaluation of MLANA and INHBA expression is as capable as one based on a combined analysis performed with genes for stemness, EMT and invasion/proliferation, in identifying melanoma subtypes that differ in their sensitivity to molecularly targeted drugs. Using this approach, we find that 75% of melanoma cell lines can be treated with either the MEK inhibitor AZD6244 or the HSP90 inhibitor 17AAG.

Received 15th August 2016.  
Accepted 11th October 2016

DOI: 10.1039/c6md00466k

www.rsc.org/medchemcomm

### Introduction

Melanoma is a heterogeneous tumor, as can be defined either by mutations or by transcriptomic analyses. More than 80% of all melanoma cells harbor a mutated gene within the MAPK pathway, all which are mutually exclusive.<sup>1,2</sup> Patients with BRAF mutations are candidates for Vemurafenib, a selective BRAF inhibitor,<sup>3</sup> and the administration of a MEK inhibitor (Trametinib) in these patients has shown to increase survival time,<sup>4</sup> demonstrating the critical function of the MAPK pathway. In addition, close to 60% of melanomas contain mutations within the CDKN2A pathway-related genes. Although these gene mutations are also mutually exclusive within the pathway, they overlap with MAPK-related mutations, indicating the presence of at least two distinct pathways that are deregulated in this tumor type. However, the mutational landscape as defined by these and other mutations has been reported not to correlate with phenotypes that are based on the transcriptome of the same tumors.<sup>1,2</sup> That gene expression-based phenotypes reveal the existence of heterogeneity among melanoma tumors is a well-established fact.<sup>1,2,5–7</sup>

Current data indicates the existence of at least 3 melanoma sub-groups as can be defined by transcriptomic analyses, with a preferential up-regulation of either differentiation, proliferation, or immune-related genes.<sup>1,2</sup> Personalized treatment, such as CTLA4 and PD-1 blockade, a most promising treatment for advanced melanoma,<sup>8</sup> is likely to benefit from these classifications as they might help design such therapy on an individual basis.<sup>9</sup> However, melanoma cells *per se* show transcriptome-based heterogeneity when studied *in vitro* as well, and this is important as it has been shown to correlate with drug sensitivity.<sup>10</sup> Thus, complete cure of melanoma might have to be pursued from two directions, one based on *in vivo* effects, and therefore the microenvironment of the tumor cells, and the other based on *in vitro* defined mechanisms related to drug sensitivity. Switching between phenotypes has been observed for melanoma *in vivo*, as well as *in vitro*, and can result in altered drug sensitivity, irrespective of the mutational status of the melanoma cells,<sup>5,11</sup> suggesting that a combination treatment consisting of two drugs affecting each one of the two phenotypes might be ultimately necessary for efficient tumor eradication. Drug resistance has traditionally been associated with a stem-cell-like phenotype and a more mesenchymal state, as compared to an epithelial one in almost all tumors.<sup>12,13</sup> The “stemness” of melanoma cells is thought to exist in a dynamic state, as melanoma cells with stem-cell markers have been reported to arise from cells lacking these markers, stem-cell like melanoma cells might not be required for tumor initiation, and the stem- and non-stem-cell phenotypes can switch dynamically.<sup>7,14</sup> Similar to the

<sup>a</sup> Department of Molecular Biology and Genetics, Bilkent University, 06800 Ankara, Turkey. E-mail: agure@bilkent.edu.tr

<sup>b</sup> Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy

<sup>c</sup> Sharett Institute of Oncology, Hadassah Hebrew University Hospital, Ein Karem Campus, 91120 Jerusalem, Israel

<sup>d</sup> Department of Pharmacy, University of Genoa, 16132 Genoa, Italy

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6md00466k

plasticity reported for stemness in melanoma, epithelial-to-mesenchymal transition (EMT) has also been implied as a factor driving metastasis and causing drug resistance in melanoma.<sup>15,16</sup> Thirdly, melanoma specific *in vitro* analyses that have aimed to explain metastatic potential have generated gene lists that could robustly classify melanoma cell lines into either a “proliferative” or an “invasive” character;<sup>6,17,18</sup> switching among these phenotypes was also shown.<sup>5,14</sup> Metastasis, drug resistance and invasion have all been attributed both to the stem cell phenotype and to EMT and have been observed in those classes defined by various researchers for melanoma. In this study, we asked if there was overlap between melanoma cells classified by these various signatures and if such melanoma subtypes would have unique responses to targeted therapy. We also aimed to identify a minimum number of biomarkers that could help classify melanoma subtypes with differing drug sensitivity profiles.

## Materials and methods

### Cell lines and tissue culture methods

Primary melanoma cell lines used in this study were previously described.<sup>19,20</sup> All cell lines (primary and commercial) were incubated in humidified incubators supplied with 5% CO<sub>2</sub> and cultured in RPMI1640 medium supplemented with 10% FBS, 1% 10 K/10 K penicillin-streptomycin (P/S Solution, Lonza, Basel, Switzerland) and 1% 200 mM L-glutamine (L-Glu, Lonza, Basel, Switzerland). Two exceptions were commercial cell lines Mewo and SK-MEL-28, which were cultured in DMEM with the same supplements. Melanocytic progeny of primary melanoma cell lines (PrMCLs) were investigated by staining for S100, MART-1, gp100 and GD2 expression using polyclonal rabbit anti-S100, monoclonal A-103, HMB-45 (Invitrogen, Carlsbad, CA, USA) and R-24 (a kind gift of Dr. Aliza Adler) antibodies, and flow cytometry-based (FACS) detection of HMW-MMA (CSPG4) (R&D Systems Minneapolis, MN, USA) (Table S1†). Samples were analyzed by FACS on a Becton Dickinson LSR II cytometer. Data analysis was performed using FCS express V.3 (De Novo Software).

### Expression profiling of INHBA and MLANA by q-RT-PCR

Quantitative profiling of INHBA and MLANA by q-RT-PCR was done using a 7500 Real-Time PCR System (Applied Biosystems, Carlsbad, CA, USA) and SYBR® Green PCR Master Mix (catalog number: 4309155, Applied Biosystems, Carlsbad, CA). PCR reactions were run under cycling conditions according to the manufacturer's instructions. GAPDH was used as the endogenous control in all reactions. PCR primers were designed using the NCBI Primer-Blast Tool (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Primer sequences were: MLANA\_F1-ATTAAGGAAGGTGCTGTGCC, MLANA\_R1-CAGCCGTGGTGAAGAGTGG, INHBA\_F1-GTTTGCCGAGTCAGGAACAGC, INHBA\_R1-GAAGAGCCAGACTTCTGCACG, GAPDH\_F1-AAAGGGTCATCATCTCTGCCCCC and GAPDH\_R1-GGCATTGCTGATGATCTTGAGGCTG. Relative expression values of INHBA

and MLANA were calculated by the ddCt method as previously described.<sup>21</sup>

### *In vitro* drug screening analyses

AZD0530, 17AAG and AZD6244 were purchased from Selleck Chemicals. Drugs were screened on transparent 96-well-plates in which cancer cell lines were seeded at 1000 cells per well 96 hours before cytotoxicity measurements. Drugs were screened using 10 different doses ranging from 0.001  $\mu$ M to 50  $\mu$ M, except 10a for which 6 different doses ranging from 0.625  $\mu$ M to 20  $\mu$ M were tested. All stock solutions were prepared in 0.1% DMSO. IC<sub>50</sub> values were calculated as previously described.<sup>22</sup> Cytotoxicity was measured with CellTiter-Glo® Luminescent Cell Viability Assay (Promega, Madison WI, USA) after 72 hours of drug treatment.

### *In silico* drug screening analyses

The normalized activity area or IC<sub>50</sub> values of 24 different commercial drugs were downloaded from the CCLE database (<https://portals.broadinstitute.org/ccle/home>) for *in silico* drug screening analyses.

### EMT, CSC and PRO/INT/INV score calculation

EMT and CSC scores were calculated based on an R script<sup>23</sup> (<http://www.R-project.org/>) which runs a principal component analysis (PCA) with the following probeset expression values: EMT scoring was based on probeset expression values of E-CAD (201131\_s\_at) and VIM (201426\_s\_at); CSC scoring used CD271 (205858\_at), CD20 (228599\_at, 228592\_at, 217418\_x\_at, 210356\_x\_at), CD133 (204304\_s\_at) and ALDH1 (212224\_at), all of which were previously published melanoma stem-cell markers.<sup>24</sup> CCLE and primary melanoma cell lines were classified as PRO (proliferative), INT (intermediate) and INV (invasive) according to Pro (*r*) and Inv (*r*) values generated using the online HOPP tool and the associated gene list.<sup>18</sup> CCLE cell lines with a Pro (*r*) value of more than 0.652 and an Inv (*r*) value of less than 0.171 were designated as “PRO”, those with a Pro (*r*) value between 0.041 to 0.630 and an Inv (*r*) value between 0.254 to 0.600 were designated as “INT” and cells with a Pro (*r*) value of less than −0.200 and an Inv (*r*) value of more than 0.445 were designated as “INV” cells. For primary cell lines, designations were as follows: PRO: Pro (*r*) > 0.533 and Inv (*r*) < 0.162; INT: Pro (*r*): 0.186 to 0.588 and Inv (*r*): 0.221 to 0.581; INV: Pro (*r*) < 0.03 and Inv (*r*) > 0.554.

### Melanoma gene expression signature analyses

E-MTAB-783 and GSE36133 commercial cancer cell line datasets were normalized separately using RMA normalization without background correction. Differentially expressed probesets were determined between melanoma cell lines and other solid tumor cell lines separately in both datasets according to *p*-value and fold change (FC) cutoffs, 0.0001 (Bonferroni corrected) and 4 (in either direction),

respectively. Significant probesets common to both datasets were determined using Venny, “Oliveros, J. C. (2007–15), an interactive tool for comparing lists with Venn diagrams (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>).” These 223 probesets constitute “gene list #1”. Standardized expression values of these probesets obtained from the C-MAP microarray data of primary melanoma cell lines<sup>20</sup> were used for hierarchical clustering analysis using Euclidean distance as the similarity metric and complete linkage as the clustering method in Cluster3.0 software. Results were visualized in JavaTreeView. Two main clusters were generated with this approach. One of the clusters showed upregulation of genes which were also upregulated in commercial melanoma cell lines compared with other solid cancer cell lines. The second cluster of primary cell lines showed upregulation of genes that were downregulated in commercial melanoma cell lines, when compared to other solid cancer cell lines. In the second round of analyses, the most differentially expressed genes between two main clusters were determined using the same significance cutoffs used in generating “gene list#1”. This resulted in “gene list#2”. INHBA and MLANA were selected among the most significantly differentially expressed genes from gene list#2 for further validation of phenotypic features of the two main clusters.

## Results & discussion

### 1. Characterizing melanoma cell lines established from primary tumor tissue

Melanoma cell lines derived from primary tumor tissues were used throughout this study. Cells were initially tested for melanoma markers S100, MART-1 and gp100 to differentiate them from non-melanoma cells.<sup>20</sup> As a further measure of specificity, gene expression data obtained from all cell lines via C-MAP technology was used to generate two gene lists representing “melanoma” (MEL) and “fibroblast” (FIB) like qualities using a previously defined algorithm.<sup>18</sup> Gene expression-based correlation measures obtained for melanoma cell lines, as well as for normal skin fibroblasts and cancer-associated fibroblasts, show that this approach is an effective means for separating melanoma from fibroblasts. As shown in Fig. 1, gene expression-based analysis revealed that only 4 out of 36 melanoma cell lines show an inverse correlation with the MEL signature, while only 6 out of 65 fibroblasts have a FIBr value lower than 0.5. The 4 melanoma cell lines that can be thus designated “fibroblast-like” are, however, clearly melanoma cell lines as judged by morphology. Other primary cell lines which were negative for S100, MART-1 or gp100, and which lacked a melanoma-like morphology, were also identified as non-melanoma and fibroblast-like by this analysis and were excluded from our studies (not shown). Similar to our analyses, “fibroblast-like” cells, as defined by low MELr and high FIBr values, are present in two out of three datasets generated from melanoma primary cell lines (Fig. S1a†). We consistently observed that such cells grow much slower, compared to those with high MELr

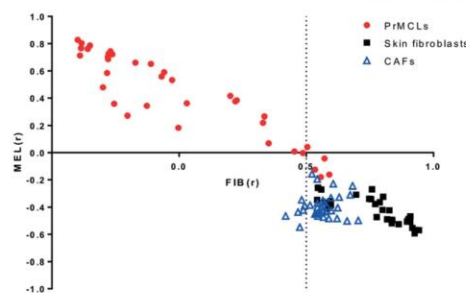


Fig. 1 Fibroblast/melanoma correlation plot of primary melanoma cell lines (PrMCLs), skin fibroblasts and cancer-associated fibroblasts (CAFs). C-MAP-based gene expression data for PrMCLs and publicly available data from skin fibroblasts (GSE11919, GSE16715, GSE21648, GSE34309, GSE35551) and CAFs (GSE46824) were used to calculate gene expression-based correlations to a fibroblast (FIB) and a melanoma (MEL) gene list. Correlation measures (Pearson's  $r$ ) for each sample are shown for FIB on the x-axis and INVMEL on the y-axis (see Materials and methods for details).

scores, explaining why they are extremely rare among commercially available melanoma cell lines (CMCL, Fig. S1b†). The average daily proliferation rate of cells with low INVMELr scores ( $<0.4$ ) was 18%, compared to 65% for cells with high INVMELr scores ( $>0.4$ ); similarly, cells with low FIBr scores ( $<0$ ), had an average proliferation rate of 67%, compared to 20% for those with high FIBr scores ( $>0$ ).

### 2. Classification of melanoma cell lines based on a combination of three gene signatures and their relation to drug sensitivity

Drug resistance in melanoma has been reported to associate with qualities determined by phenotypes such as EMT, stemness, and measures pertaining to the cell's proliferative and invasive capacity. We evaluated each primary melanoma cell line, as well as CMCLs, for each of these 3 qualities *in silico*, to determine any overlap between these definitions, and if one or a combination of these could be used to better explain sensitivity to a molecularly targeted drug. For this purpose, cells were first classified into proliferative (PRO), invasive (INV) or intermediate (INT) categories using an online algorithm based on a previously published gene list.<sup>18</sup> Each cell was then assigned a score corresponding to its epithelial/mesenchymal state (EMT score), in addition to a score reflecting how stem-like the cell was (CSC score; see methods). Not surprisingly, only one CMCL and 3 PrMCLs had both epithelial and CSC-like qualities, as these two phenotypes are expected to show mutual exclusivity (quadrant 1, Fig. 2). Similarly, most INV cells were more mesenchymal than epithelial, while PRO cells were more epithelial. However, while the INT group among CMCLs was mostly of a CSC-like and mesenchymal phenotype (Q3, Fig. 2, top), the INT group among primary melanoma cell lines had more

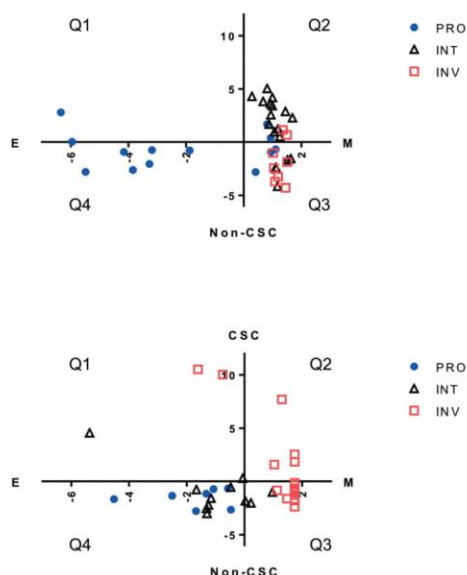


Fig. 2 Clustering melanoma cell lines based on different molecular signatures. Distribution of CMCLs (top) and PrMCLs (bottom), categorized into PRO, INV or INT groups according to an EMT score (X axis) and stemness score (Y axis). E: epithelial, M: mesenchymal phenotype. CSC: cancer stem-cell-like phenotype, non-CSC: non-cancer-stem-cell-like phenotype. Cells are shown distributed into 4 quadrants: Q1: E and CSC-like, Q2: M and CSC-like, Q3: M and non-CSC-like, and Q4: E and non-CSC-like.

non-CSC-like and epithelial features (Q4, Fig. 2, bottom). We analyzed if cells harboring BRAF (V600) mutations were preferentially of a particular transcriptomic subtype. CSC and EMT scores were not significantly different between mutant and WT cells. However, the PRO scores of mutant cells were significantly larger than that of WT cells, and the inverse was observed for INV scores;  $p < 0.01$  and  $p = 0.03$ , respectively. This suggested that BRAF mutations were not a primary determinant of stemness or the EMT state of cells, but were related to PRO and INV scores.

We then asked if the classification based on the combination of these three transcriptomic groups could predict drug sensitivity better than that which could be predicted by using the "PRO/INT/INV" classification by itself. For this purpose, we correlated INV and PRO values obtained for melanoma cell lines for which there was cytotoxicity data in the CCLE database.<sup>22</sup> The strongest direct correlation between PRO values and simultaneous negative correlation with INV values was observed for cytotoxicity data from two MEK inhibitors, AZD6244 and PD0325901, in line with the correlation of these phenotypes with BRAF mutation status. The exact opposite pattern was observed most prominently for three other drugs, the HSP90 inhibitor 17-AAG, the SRC-inhibitor AZD0530, and

a cMET inhibitor PF-2341066 (Table S2†). As HSP90 and MEK inhibitors have previously been shown to affect opposite phenotypes,<sup>25</sup> and similar findings exist for SRC and MEK inhibitor combinations,<sup>10</sup> we decided to focus on 17AAG, AZD6244 and AZD0530. We first determined *in silico* how melanoma cells, as defined by the combination of 3 scores, behaved in response to these drugs and whether this classification would help improve on the PRO/INT/INV classification. As shown in Fig. 3, PRO cells were primarily sensitive to the MEK inhibitor AZD6244, while INV cells were to 17AAG (HSP90 inhibitor) and AZD0530 (SRC inhibitor). By further classifying cells into quadrants based on the added EMT and stemness score, we were unable to identify a subpopulation that was more sensitive to any of the three drugs tested, with the most resistant and sensitive groups identified based on quarters showing similar values to those obtained for PRO, INT and INV groups (Fig. 3). This indicates that classifying cells based on a combination of EMT and CSC markers is able to identify distinct cell populations that differ in their chemosensitivity profiles, but does not help identify cells that are more sensitive to these drugs, compared to the classification generated by the PRO/INV score.

Next, we wanted to validate the *in silico* cytotoxicity correlations *in vitro* using PrMCLs. In contrast to what we observed for CMCLs, PRO, INT and INV groups were not as clearly separated into quadrants in analyses of PrMCLs (Fig. 2). Most PRO cells were observed in Q4 (epithelial and non-CSC-like), together with INT cells, while INV cells were present in quarters 1, 2 and 3, which is a different distribution compared to the classification we observed for CMCLs. Cytotoxicity profiles of PrMCLs, however, were quite different between Q-based and PRO/INV-based classes (Fig. 4). For 17-AAG, although INV cells were the most sensitive, INT cells were generally more resistant than PRO cells. A single cell tested in Q2 was the most sensitive, supporting this observation. For AZD6244, PRO and Q4 cells were relatively sensitive. Although a single cell in Q2 was also sensitive, as this is a sole observation, we do not think it indicates an exception. AZD0530, similar to its *in silico* data, did not show much variation among groups, although INV cells and that in Q2 were somewhat more sensitive. We then tested a novel drug 10a, that has been shown to be a specific SRC inhibitor.<sup>26</sup> 10a was effective in particular on PRO cells and those in Q4 (epithelial, non-CSC-like), but also on cells in Q2 (mesenchymal, CSC-like) (Fig. 4). There were no statistically significant differences among either quadrant or PRO/INV-based classes for 17AAG and AZD0530 IC<sub>50</sub> values. However, there were significant differences for AZD6244 between Q3 vs. Q4 ( $p = 0.03$ , *t*-test), and for 10a, in quadrant based ( $p = 0.026$ ) or PRO/INT/INV based ( $p < 0.001$ ) classifications as identified by 1-way ANOVA. This observation reveals that PRO/INV-based and quadrant-based classifications are distinct, in that the cells categorized by both differ in their responses to targeted therapy, especially to the SRC inhibitor 10a. Interestingly, the sensitivity profile of 10a is the complete inverse of that observed for AZD0530. Our data shows that SRC inhibition by

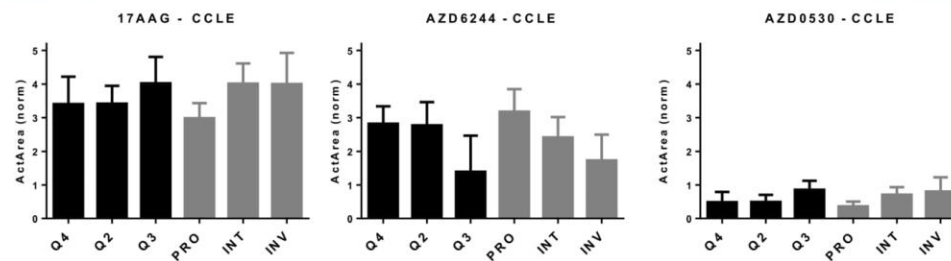


Fig. 3 *In silico* cytotoxicity analyses of molecularly targeted drugs on CMCLs. CMCLs clustered in quadrants 2, 3, and 4 (as indicated in Fig. 2) as well as according to the PRO/INV classification were analyzed for sensitivity against the shown drugs. Cytotoxicity is shown as normalized activity area (Y axis). 1-Way ANOVA-based analysis of quadrant-based classification of cells gave *p* values of 0.07, 0.034 and 0.05 for 17AAG, AZD6244 and AZD0530, respectively, while the same statistical analysis for PRO, INT and INV groups generated *p* values of 0.025, 0.023 and 0.03 for 17AAG, AZD6244 and AZD0530, respectively. Q2: M and CSC-like, Q3: M and non-CSC-like, and Q4: E and non-CSC-like.

10a exhibits a profile almost identical to that of AZD6244 for PrMCLs (Fig. 4). As MEK and SRC are shared in most pathways, this suggests the SRC-MAPK pathway is active and responsible for growth primarily in PRO cells and those in quarters 2 and 4. AZD0530 shows minimal activity in both CMCLs and PrMCLs at the concentrations tested and does not happen to have distinctly different effects on subgroups of melanoma. That its activity profile is different from 10a suggests either that AZD0530 is unable to inhibit SRC in these cells or that 10a is inhibiting a non-SRC protein.

We also compared cytotoxicity responses among cells with and without BRAF mutations. We did not observe any significant differences in  $IC_{50}$  values for any of the drugs tested between the two groups, although BRAF mutant cells were gen-

erally more sensitive to the MEK inhibitor AZD6244 (data not shown).

### 3. MLANA and INHBA as biomarkers of molecularly targeted drug sensitivity in melanoma

All phenotypical categorizations described thus far are based on the evaluation of a relatively large number of genes. However, as the various categories defined by the three different phenotypical markers generate clusters with similar drug sensitivity profiles, we argued that such profiles could be reflected by a much smaller number of gene expression patterns. As a number of genes used in these classifiers are related to differentiation markers of melanoma, as a discovery strategy, we sought to identify genes whose expression could differentiate melanoma from non-melanoma tumors by comparing gene expression profiles of tumors within E-MTAB-783 and GSE36133 datasets. 223 probesets with a Bonferroni corrected *p* value of 0.0001 or below and a difference of more than 4 fold that were common to both datasets were used to cluster PrMCLs *in silico* (Fig. S2†). This analysis revealed two distinct clusters (A and B) which were then used to discover genes that were most differentially expressed between these two clusters (Fig. S2†). MLANA and INHBA were selected as representatives of the most upregulated genes in each of the two distinct clusters, especially since they have been reported to identify opposite phenotypes in melanoma: MLANA is critical for the regulation of mammalian pigmentation,<sup>27</sup> while INHBA is reported to decrease melanin content of B16 murine melanoma cell lines.<sup>28</sup> Furthermore, INHBA is a member of the transforming growth factor superfamily which induces a mesenchymal phenotype by increasing N-cadherin expression in esophageal carcinoma cell lines.<sup>29</sup> These two genes are inversely expressed in both commercial and PrMCLs (Fig. S3†). Although most INHBA expressing cells are low expressors of MLANA and *vice versa* (thus in quadrants 1 and 3, Fig. S4†), there are cells that express both or neither. Both genes have been reported to show a strong correlation between mRNA and protein levels;<sup>17,30</sup> we therefore believe the

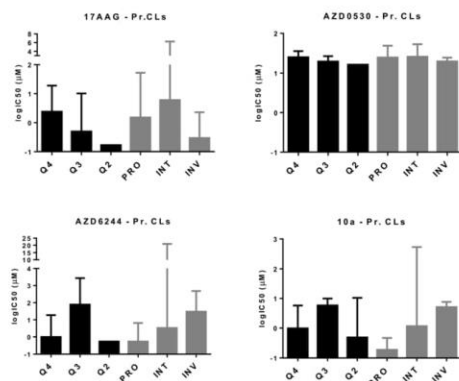


Fig. 4 *In vitro* cytotoxicity analyses of molecularly targeted drugs on melanoma cell lines. CMCLs and PrMCLs clustered in quadrants 2, 3, and 4 (Fig. 2) as well as according to the PRO/INV classification were analyzed for sensitivity against the shown drugs. Cytotoxicity is shown as log  $IC_{50}$  (Y axis). Q2: M and CSC-like, Q3: M and non-CSC-like, and Q4: E and non-CSC-like.

mRNA-based phenotype should be reflected at the protein level as well. As expected, MLANA and INHBA mRNA expression correlated tightly with MITF and AXL, two genes that are known to play critical roles in phenotype switching in melanoma<sup>31,32</sup> (Fig. S5†). In line with BRAF mutations not correlating with previously reported transcriptomic signatures for melanoma, the BRAF mutation status of cells does not seem to correlate with MLANA or INHBA either (Fig. S4†).

We observed that MLANA and INHBA expression-based classification was successful in revealing drug sensitivity differences for both CMCLs and PrMCLs: we combined cells in Q2, 3 and 4 as their sensitivity profiles were similar. We then compared their sensitivity to cells within Q1 (MLANA-high, INHBA-low). This revealed the expected trends for all three drugs. *In silico* analysis showed that Q1 cells (MLANA-high, INHBA-low) were relatively sensitive to AZD6244 (MEK inhibitor), while they were relatively resistant to 17AAG (HSP90 inhibitor) and AZD0530 (SRC inhibitor), compared to cells in Q2, 3, 4 (Fig. 5). We then validated these findings *in vitro* for PrMCLs and included the SRC inhibitor 10a in our analysis. As shown in Fig. 6, AZD6244 and 10a are both more effective on Q1 (MLANA-high, INHBA-low) cells, while 17AAG and AZD0530 showed the exact opposite trend. The differential effect of AZD6244 was statistically significant. To test if the cells that were resistant to AZD6244 were also sensitive to 17AAG or 10a, we plotted the drug sensitivity profiles of cells for either of the two drugs. This analysis, however, revealed that while half of either CMCLs or PrMCLs followed this pattern, about a quarter were sensitive to both, and another quarter of the cells were resistant to both drugs analyzed (Fig. S6†). As previously observed, the sensitivity profiles of 10a and AZD6244 were highly similar (Fig. S6c†). We therefore conclude that phenotyping based on MLANA and INHBA will be able to identify cells that are sensitive to either AZD6244/10a or 17AAG, in about 50% of the cells tested. 25% of cells will be sensitive to both drugs, and a quarter might be resistant to both.

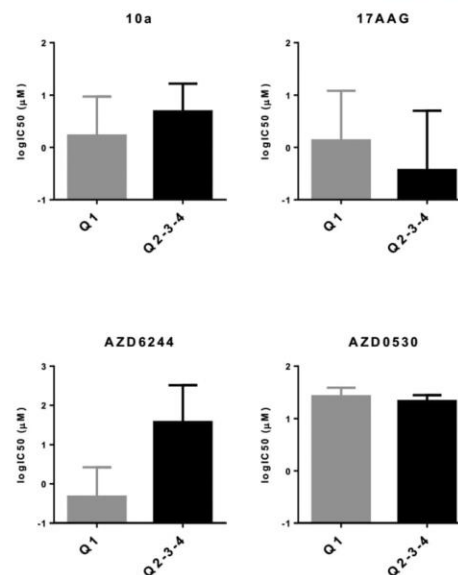


Fig. 6 MLANA-INHBA-based drug sensitivity analysis of PrMCLs. *In vitro* cytotoxicity analysis of PrMCLs against 4 molecularly targeted drugs. Cells contained within quadrants identified in Fig. S3† are indicated. *P* values (*t*-test) for 10a, 17AAG, AZD6244, and AZD0530 are 0.1, 0.43, 0.01 and 0.32, respectively. Q1: E and CSC-like, Q2: M and CSC-like, Q3: M and non-CSC-like, and Q4: E and non-CSC-like.

## Discussion

Plasticity of stemness, EMT and INV/PRO phenotypes in melanoma are clearly related to changes in the signaling pathway choices of these cells. We hypothesized that different melanoma cells, as characterized by gene expression profiles

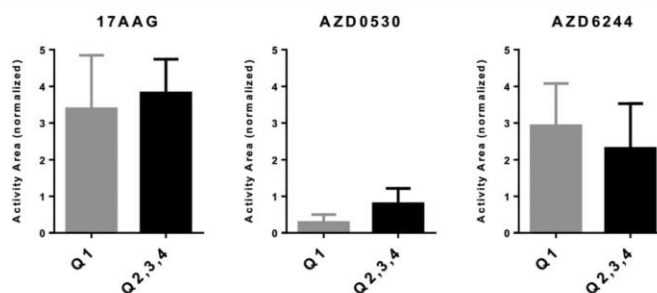


Fig. 5 MLANA-INHBA-based drug sensitivity analysis of CMCLs. *In silico* cytotoxicity analysis of CMCLs against 3 molecularly targeted drugs. Cells contained within quadrants identified in Fig. S3† are indicated. *P* values (*t*-test) for 17AAG, AZD0530 and AZD6244 are 0.26, 0.0002 and 0.13, respectively. Q1: E and CSC-like, Q2: M and CSC-like, Q3: M and non-CSC-like, and Q4: E and non-CSC-like.

reported to relate to distinct phenotypes, would primarily rely on one particular signaling pathway which could be blocked by one of the drugs generated to target specific signaling pathways. We also hypothesized that the various phenotypes defined for melanoma could overlap. Despite the limited number of cell lines used here, we have been able to support our second hypothesis, as we were able to show that cells grouped by utilizing multiple gene lists are not necessarily better characterized in terms of their drug sensitivity profiles, compared to using any given one. If each of these different phenotypes arises due to the utilization of a distinct pathway, then the phenotypic switch should relate to a shift in the signaling pathways the cell uses, resulting in altered drug sensitivity to a molecularly targeted drug. In line with this argument, we were able to generate partial support for our first hypothesis: although we were able to identify distinct drugs that were more effective on cells as defined by a given phenotype, and that drugs that target different molecules within the same pathway show similar outcomes (shown for 10a and AZD6244), we observed that drugs showing opposite correlations with a given phenotype (AZD6244 or 10a *versus* 17AAG) did not necessarily have mutually exclusive actions on cells. This is in contrast to the fact that the biomarkers identifying these cells are almost mutually exclusive (MLANA and INHBA). The explanation for this could be that mRNA-based biomarkers, however good they may be in identifying cells with homogenous transcriptomic phenotypes, are not necessarily identifying cells that utilize the same survival pathways. This discrepancy could then likely be due to the fact that simultaneous but distinct pathways are active in a given cell and they have to be targeted at the same time for effective control of cell growth. Despite not being statistically significant, the similarity in sensitivity profiles we observe for 17AAG and AZD0530, two drugs that target non-overlapping signal pathways, support this line of thought. It is possible that the ultimate aim of the cell is to alter its transcriptomic profile, generating a given functional phenotype. If various pathways can be utilized for this purpose, then the cell will use one out of the several ones possible. On the other hand, the fact that melanoma cells that develop resistance to BRAF inhibitors re-activate the MAPK pathway suggests there are situations where the cell relies solely on one pathway.<sup>33</sup>

Targeted therapy *via* agents like Dabrafenib (BRAF inhibitor) and Trametinib (MEK inhibitor) significantly increase overall survival of metastatic melanoma patients with BRAF V600E or V600K mutations.<sup>34,35</sup> That MEK and RAF inhibitor cytotoxicity profiles are similar according to our *in silico* analysis is also supportive of this. One reason we could not correlate BRAF mutations to sensitivity to the drugs tested here could be the limited number of mutant cells we studied. This is supported by the fact that BRAF V600 mutant cells have much higher PRO scores compared to INV scores, and the fact that PRO/INV scores do correlate with drug sensitivity, especially to that of the MEK inhibitor AZD6244. However, melanoma subclasses as defined by transcriptomic analyses cannot be explained based on their mutational profile.<sup>1,2</sup> It is

therefore likely that a simple classification based on MLANA and INHBA can help identify tumors that will benefit from one of the drugs analyzed in this study more by itself or in combination with mutation information.

## Conclusions

Our results indicate that melanoma cells are possibly utilizing multiple survival pathways and that the best personalized treatment approach will be one that first reveals the signaling pathways critical for the survival in each cell, followed by the utilization of a drug or a combination of drugs that can target all of these pathways at once. Our data shows that a profiling based on gene expression testing of only MLANA and INHBA is as capable of classifying *in vitro* responses to molecularly targeted drugs. Larger validation experiments *in vivo* and tests of drug synergy will need to be performed to confirm this observation.

## Abbreviations

|       |                                      |
|-------|--------------------------------------|
| CCLE  | Cancer Cell Line Encyclopedia        |
| C-MAP | Connectivity map                     |
| CMCL  | Commercial melanoma cell line        |
| CSC   | Cancer stem cell                     |
| EMT   | Epithelial to mesenchymal transition |
| FIB   | Fibroblast                           |
| INHBA | Inhibin beta A                       |
| INT   | Intermediate                         |
| INV   | Invasive                             |
| MEL   | Melanoma                             |
| MLANA | Melan-A                              |
| PrMCL | Primary melanoma cell line           |
| PRO   | Proliferative                        |
| Q     | Quadrant                             |
| RMA   | Robust multi-array average           |

## References

- 1 N. Cancer Genome Atlas, *Cell*, 2015, **161**, 1681–1696.
- 2 H. Cirenajwis, H. Ekedahl, M. Lauss, K. Harbst, A. Carneiro, J. Enoksson, F. Rosengren, L. Werner-Hartman, T. Torngren, A. Kvist, E. Fredlund, P. O. Bendahl, K. Jirstrom, L. Lundgren, J. Howlin, A. Borg, S. K. Gruvberger-Saal, L. H. Saal, K. Nielsen, M. Ringner, H. Tsao, H. Olsson, C. Ingvar, J. Staaf and G. Jonsson, *Oncotarget*, 2015, **6**, 12297–12309.
- 3 P. B. Chapman, A. Hauschild, C. Robert, J. B. Haanen, P. Ascierto, J. Larkin, R. Dummer, C. Garbe, A. Testori, M. Maio, D. Hogg, P. Lorigan, C. Lebbe, T. Jouary, D. Schadendorf, A. Ribas, S. J. O'Day, J. A. Sosman, J. M. Kirkwood, A. M. Eggermont, B. Dreno, K. Nolop, J. Li, B. Nelson, J. Hou, R. J. Lee, K. T. Flaherty, G. A. McArthur and B.-S. Group, *N. Engl. J. Med.*, 2011, **364**, 2507–2516.
- 4 K. T. Flaherty, C. Robert, P. Hersey, P. Nathan, C. Garbe, M. Milhem, L. V. Demidov, J. C. Hassel, P. Rutkowski, P. Mohr, R. Dummer, U. Trefzer, J. M. Larkin, J. Utikal, B. Dreno, M. Nyakas, M. R. Middleton, J. C. Becker, M. Casey, L. J.

- Sherman, F. S. Wu, D. Ouellet, A. M. Martin, K. Patel, D. Schadendorf and M. S. Group, *N. Engl. J. Med.*, 2012, **367**, 107–114.
- 5 K. S. Hoek, O. M. Eichhoff, N. C. Schlegel, U. Dobbeling, N. Kobert, L. Schaerer, S. Hemmi and R. Dummer, *Cancer Res.*, 2008, **68**, 650–656.
- 6 A. R. Jeffs, A. C. Glover, L. J. Slobbe, L. Wang, S. He, J. A. Hazlett, A. Awasthi, A. G. Woolley, E. S. Marshall, W. R. Joseph, C. G. Print, B. C. Baguley and M. R. Eccles, *PLoS One*, 2009, **4**, e8461.
- 7 A. Roesch, M. Fukunaga-Kalabis, E. C. Schmidt, S. E. Zabierowski, P. A. Brafford, A. Vultur, D. Basu, P. Gimotty, T. Vogt and M. Herlyn, *Cell*, 2010, **141**, 583–594.
- 8 J. D. Wolchok, H. Kluger, M. K. Callahan, M. A. Postow, N. A. Rizvi, A. M. Lesokhin, N. H. Segal, C. E. Ariyan, R. A. Gordon, K. Reed, M. M. Burke, A. Caldwell, S. A. Kronenberg, B. U. Agunwamba, X. Zhang, I. Lowy, H. D. Inzunza, W. Feely, C. E. Horak, Q. Hong, A. J. Korman, J. M. Wigginton, A. Gupta and M. Sznol, *N. Engl. J. Med.*, 2013, **369**, 122–133.
- 9 M. Lauss, J. Nsengimana, J. Staaf, J. Newton-Bishop and G. Jönsson, *J. Invest. Dermatol.*, 2016, DOI: 10.1016/j.jid.2016.05.119, in press.
- 10 J. Ferguson, I. Arozarena, M. Ehrhardt and C. Wellbrock, *Oncogene*, 2013, **32**, 86–96.
- 11 M. Saez-Ayala, M. F. Montenegro, L. Sanchez-Del-Campo, M. P. Fernandez-Perez, S. Chazarra, R. Freter, M. Middleton, A. Pinero-Madrona, J. Cabezas-Herrera, C. R. Goding and J. N. Rodriguez-Lopez, *Cancer Cell*, 2013, **24**, 105–119.
- 12 A. Singh and J. Settleman, *Oncogene*, 2010, **29**, 4741–4751.
- 13 A. Mitra, L. Mishra and S. Li, *Oncotarget*, 2015, **6**, 10697–10711.
- 14 K. S. Hoek and C. R. Goding, *Pigm. Cell Melanoma Res.*, 2010, **23**, 746–759.
- 15 A. Jayachandran, M. Anaka, P. Prithviraj, C. Hudson, S. J. McKeown, P. H. Lo, L. J. Vella, C. R. Goding, J. Cebon and A. Behren, *Oncotarget*, 2014, **5**, 5782–5797.
- 16 B. Kreiseder, Y. M. Holper-Schichl, B. Muellauer, N. Jacobi, A. Pretsch, J. A. Schmid, R. de Martin, H. Hundsberger, A. Eger and C. Wiesner, *PLoS One*, 2015, **10**, e0119402.
- 17 K. S. Hoek, N. C. Schlegel, P. Brafford, A. Sucker, S. Ugurel, R. Kumar, B. L. Weber, K. L. Nathanson, D. J. Phillips, M. Herlyn, D. Schadendorf and R. Dummer, *Pigm. Cell Res.*, 2006, **19**, 290–302.
- 18 D. S. Widmer, P. F. Cheng, O. M. Eichhoff, B. C. Belloni, M. C. Zipser, N. C. Schlegel, D. Javelaud, A. Mauviel, R. Dummer and K. S. Hoek, *Pigm. Cell Melanoma Res.*, 2012, **25**, 343–353.
- 19 M. Lotem, T. Peretz, O. Drize, Z. Gimmon, D. Ad El, R. Weitzen, H. Goldberg, I. Ben David, D. Prus, T. Hamburger and E. Shiloni, *Br. J. Cancer*, 2002, **86**, 1534–1539.
- 20 M. Lotem, S. Merims, S. Frank, T. Hamburger, A. Nissan, L. Kadouri, J. Cohen, R. Straussman, G. Eisenberg, S. Frankenburg, E. Carmon, B. Alaiyan, S. Shneibaum, Z. Ozge Ayyildiz, M. Isbilen, K. Mert Senses, I. Ron, H. Steinberg, Y. Smith, E. Shiloni, A. O. Gure and T. Peretz, *J. Immunol. Res.*, 2016, **2016**, 8121985.
- 21 K. J. Livak and T. D. Schmittgen, *Methods*, 2001, **25**, 402–408.
- 22 J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehar, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi Jr., M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palessandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel and L. A. Garraway, *Nature*, 2012, **483**, 603–607.
- 23 T. R. D. C. Team, *R: A Language and Environment for Statistical Computing*, 2010.
- 24 J. P. Medema, *Nat. Cell Biol.*, 2013, **15**, 338–344.
- 25 M. Zajac, G. Gomez, J. Benitez and B. Martinez-Delgado, *BMC Med. Genomics*, 2010, **3**, 44.
- 26 T. Kruewel, S. Schenone, M. Radi, G. Maga, A. Rohrbeck, M. Botta and J. Borlak, *PLoS One*, 2010, **5**, e14143.
- 27 T. Hoashi, H. Watabe, J. Muller, Y. Yamaguchi, W. D. Vieira and V. J. Hearing, *J. Biol. Chem.*, 2005, **280**, 14006–14016.
- 28 M. Murakami, F. Matsuzaki and M. Funaba, *Mol. Biol. Rep.*, 2008, **36**, 1247–1250.
- 29 K. Yoshinaga, H. Inoue, T. Utsunomiya, H. Sonoda, T. Masuda, K. Mimori, Y. Tanaka and M. Mori, *Clin. Cancer Res.*, 2004, **10**, 5702–5707.
- 30 J. Du, A. J. Miller, H. R. Widlund, M. A. Horstmann, S. Ramaswamy and D. E. Fisher, *Am. J. Pathol.*, 2003, **163**, 333–343.
- 31 K. Kemper, P. L. de Goeje, D. S. Peeper and R. van Amerongen, *Cancer Res.*, 2014, **74**, 5937–5941.
- 32 D. J. Konieczkowski, C. M. Johannessen, O. Abudayyeh, J. W. Kim, Z. A. Cooper, A. Piris, D. T. Frederick, M. Barzily-Rokni, R. Straussman, R. Haq, D. E. Fisher, J. P. Mesirov, W. C. Hahn, K. T. Flaherty, J. A. Wargo, P. Tamayo and L. A. Garraway, *Cancer Discovery*, 2014, **4**, 816–827.
- 33 H. Shi, W. Hugo, X. Kong, A. Hong, R. C. Koya, G. Moriceau, T. Chodon, R. Guo, D. B. Johnson, K. B. Dahlman, M. C. Kelley, R. F. Kefford, B. Chmielowski, J. A. Glaspy, J. A. Sosman, N. van Baren, G. V. Long, A. Ribas and R. S. Lo, *Cancer Discovery*, 2014, **4**, 80–93.
- 34 K. Margolin, *Curr Treat Options Oncol*, 2016, **17**, 48.
- 35 C. Robert, B. Karaszewska, J. Schachter, P. Rutkowski, A. Mackiewicz, D. Stroiakovski, M. Lichinitser, R. Dummer, F. Grange, L. Mortier, V. Chiarion-Sileni, K. Drucis, I. Krajsova, A. Hauschild, P. Lorigan, P. Wolter, G. V. Long, K. Flaherty, P. Nathan, A. Ribas, A. M. Martin, P. Sun, W. Crist, J. Legos, S. D. Rubin, S. M. Little and D. Schadendorf, *N. Engl. J. Med.*, 2015, **372**, 30–39.

## Predictive Gene Signature for Pyrazolopyrimidine Derivative c-Src Inhibitor 10a Sensitivity in Melanoma Cells

Baris Kucukkaraduman, Can Turk, Anna L. Fallacara, Murat Isbilen, Kerem M. Senses, Zeynep O. Ayyildiz, Muhammad W. Akbar, Michal Lotem, Maurizio Botta, and Ali O. Gure\*

Cite This: ACS Med. Chem. Lett. 2020, 11, 928–932

Read Online

ACCESS |

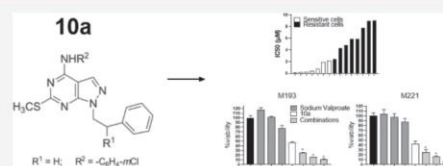
Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** Melanoma is a highly aggressive cancer with poor prognosis. Although more than 80% of melanomas harbor an activating mutation in genes within the MAPK pathway, which are mutually exclusive, usefulness of therapies targeting MAPK pathway are impeded by innate and/or acquired resistance in most patients. In this study, using melanoma cells, we report the efficacy of a recently developed pyrazolo[3,4-*d*]pyrimidine derived c-Src inhibitor 10a and identify a molecular signature which is predictive of 10a chemosensitivity. We show that the expression of TMED7, PLOD2, XRCC5, and NSUN5 are candidate biomarkers for 10a sensitivity. Although an undifferentiated/mesenchymal/invasive status of melanoma cells is associated with resistance to 10a, we show here for the first time that melanoma cells can be sensitized to 10a via treatment with valproic acid, a histone deacetylase inhibitor.

**KEYWORDS:** melanoma, c-Src, drug resistance, valproic acid, sensitivity biomarkers



Melanoma is one of the most aggressive types of cancers with heterogeneous features.<sup>1</sup> Treatment strategies and success rates vary depending on many factors, including feasibility of surgery, stage of the melanoma, presence of specific genetic mutations, and the rate of tumor growth. While surgery is the standard approach for localized and early stage melanoma, advanced stages require lymph node dissection, radiation therapy, and medication assisted therapies, including immunotherapy, chemotherapy, and targeted therapy.<sup>2,3</sup> However, long-term efficacy of such treatments is limited by their side effects and innate or acquired resistance. Therefore, new treatment strategies might help improve patient's quality of life and increase survival time.

The c-Src proto-oncogene is a nonreceptor protein kinase which is activated by different classes of cellular receptors, resulting in a diverse spectrum of biological activities including cell cycle progression, apoptosis, cell adhesion, migration, transformation, and various immune responses.<sup>4,5</sup> Advanced stages of cancer, tumor growth, progression, and metastasis are associated with increased expression and activity of c-Src in melanoma and other solid tumors.<sup>4,6</sup> Hence, c-Src represents a promising therapeutic target for the treatment of melanoma.

In this line, several Src inhibitors have been available for clinical assessment, including dasatinib and saracatinib, which are utilized in clinical trials in melanoma. Dasatinib is known to inhibit Src and other tyrosine kinases, including c-KIT, PDGFR, c-FMS, EPHA receptor, and BCR-ABL fusion.<sup>7</sup> Nevertheless, a clinical trial has suggested that daily dasatinib treatment has only a minimal benefit in unresectable

melanoma patients, the 6-month progression free survival (PFS) rate being only 13% among these patients and with high rates of toxicity.<sup>8</sup> Similarly, saracatinib was also minimally effective in a phase II study.<sup>9</sup> Hence, there is a need for more potent c-Src inhibitors and/or a better definition of melanoma subtypes that would be suitable for Src inhibitor-based treatment protocols.

Pyrazolo[3,4-*d*]pyrimidine derivatives show antitumor activity by inhibiting several serine/threonine or tyrosine kinases.<sup>10</sup> 10a is the 6-methylthio derivative of pyrazolo[3,4-*d*]pyrimidine, which was shown to inhibit c-Src selectively.<sup>11</sup> In this regard, we investigated the efficacy and molecular mechanisms of response to 10a in human melanoma cells. In this study, based on cytotoxic activity of 10a in 15 melanoma cells, we report the identification of a new gene signature which can predict chemosensitivity to 10a. Two distinct phenotypes of cells, defined as sensitive and resistant, were further analyzed to reveal an underlying mechanism for this differential response to 10a. We found that proliferative or mesenchymal features of the cells are associated with distinct sensitivity of 10a. Through a protein–chemical interaction network analysis, we identified that three histone deacetylase

Special Issue: In Memory of Maurizio Botta: His Vision of Medicinal Chemistry

Received: December 31, 2019

Accepted: February 18, 2020

Published: February 18, 2020



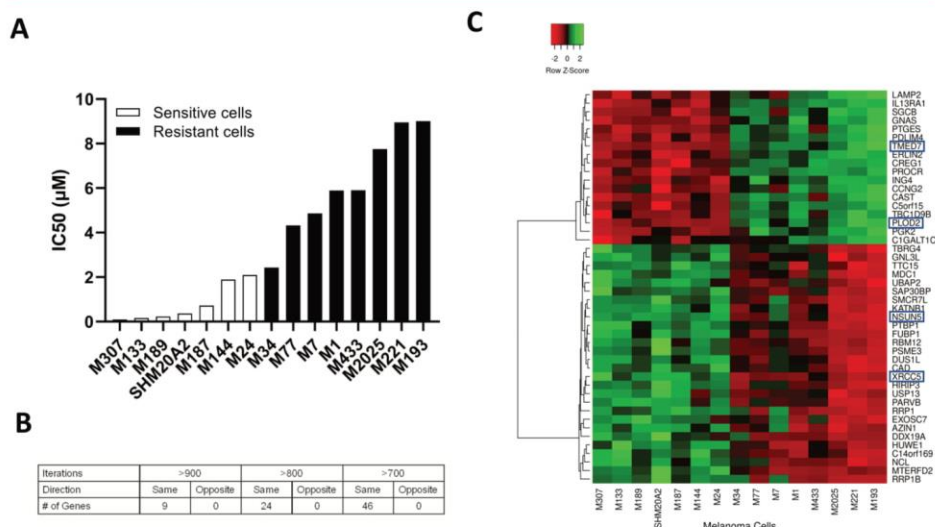


**Figure 1.** *In vitro* cytotoxicity analyses of c-Src inhibitor 10a on melanoma cells and hierarchical clustering of cell lines according to 10a chemosensitivity biomarker genes. (A) Fifteen melanoma cells were analyzed for sensitivity against 10a. Cytotoxicity is expressed in terms of IC<sub>50</sub> (y-axis). (B) 10a chemosensitivity biomarker genes were determined by 2-fold cross validation with 1000 iterations, and 9, 24, and 46 genes had a frequency of significance in more than 900, 800, and 700 iterations, respectively. (C) Melanoma cells were hierarchically clustered with 46 genes that were significantly correlated with cytotoxicity values against 10a for more than 700 iterations. Two clusters were defined as sensitive (M307, M133, M189, SHM20A2, M187, M144, M24) and resistant (M34, M77, M1, M433, M2025, M221, M193). Standardized expression values were used for clustering, and clustering was performed with the complete linkage method on the heatmap online tool.<sup>15</sup>

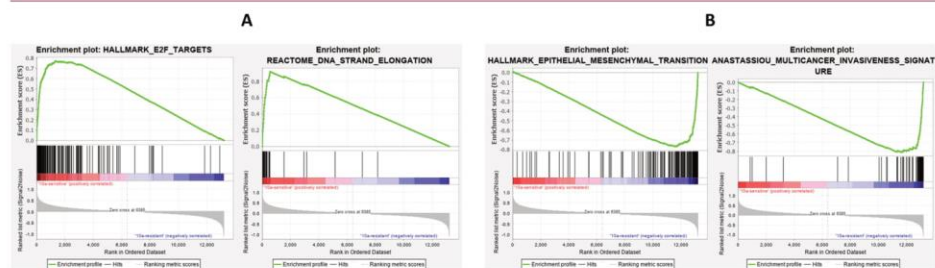


**Figure 2.** Comparative analysis of differentially enriched molecular gene sets among sensitive and resistant cells by GSEA. (A) Cells were grouped as sensitive and resistant based on clustering with the 10a chemosensitivity associated gene list seen in Figure 1C. Gene sets including cell cycle related targets of E2F transcription factor (left) and DNA strand elongation related genes (right) were enriched in 10a-sensitive melanoma cells. (B) Gene sets defining epithelial to mesenchymal transition as in wound healing, fibrosis, and metastasis (left) and multicancer invasion associated genes (right) showed enrichment in 10a-resistant melanoma cells. FDR values for each enrichment plot were less than 0.001.

inhibitors, valproic acid, entinostat, and trichostatin A, were predicted to synergize with 10a. The synergizing effect of valproic acid was validated *in vitro*. We also aimed to define a minimal number of genes that could be used as biomarkers of 10a sensitivity. We show that the expression level of four genes can be used to predict drug sensitivity against 10a.

When tested on 15 melanoma cells, cytotoxic responses to 10a occurred over a remarkably wide range of doses with IC<sub>50</sub> values ranging from 0.09 to 9.01 μM (Figure 1A, Table S1). This variation strongly suggests underlying molecular differ-

ences among these cells and thus prompted us to identify a gene signature of chemosensitivity for 10a. For this purpose, we correlated gene expression data from melanoma cells obtained by C-MAP analysis with IC<sub>50</sub> values using a 2-fold cross validation with a 1000 iteration method (see Supporting Information).<sup>12</sup> Briefly, in each iteration, cells were randomly divided into two groups, and a Pearson's *r* value between 10a IC<sub>50</sub> and gene expression data for each gene was calculated for these groups. This was repeated 1000 times. We then determined whether the correlations obtained for a given

gene were consistently in the same direction (either positively or negatively correlated in both groups) where the results were statistically significant in all groups. This analysis showed that expression of 9 genes was consistently and significantly correlated with  $IC_{50}$  values in more than 900 iterations in the same direction. In more than 800 iterations, expression levels of 24 genes showed consistent and significant correlations with  $IC_{50}$  values in the same directions. We chose 46 genes whose expression profile showed significant concordance with chemosensitivity of 10a in more than 700 iterations for classifying melanoma cells (Figure 1B). Hierarchical clustering analysis with these 46 genes revealed two subgroups of cells which are relatively sensitive and resistant to 10a (Figure 1C).

To investigate molecular mechanisms underlying the differential response to 10a, we performed gene set enrichment analysis (GSEA)<sup>13</sup> with sensitive and resistant subgroups defined with the 46-gene list. M307, M133, M189, SHM20A2, M187, M144, and M24 cells constituted the sensitive subgroup while M34, M77, M7, M1, M433, M2025, M221, and M193 constituted the resistant subgroup. GSEA revealed a significant enrichment for gene sets, including E2F target genes related to cell cycle and DNA strand elongation related genes in the sensitive subgroup, suggesting a proliferative phenotype in these cells (Figure 2A, Table S2). On the other hand, resistant subgroups showed significant enrichments of gene sets related to epithelial-mesenchymal transition and multicancer invasiveness signatures, indicating an invasive/mesenchymal phenotype in resistant cells (Figure 2B). These results are concordant with previous findings where melanoma cells were classified into invasive, proliferative, and intermediate categories.<sup>14</sup>

To further investigate mechanisms behind resistance to 10a, we performed protein–chemical interaction network analysis with 46 genes using the NetworkAnalyst<sup>16</sup> online tool. This revealed that, out of 18 resistance associated proteins, 15 extensively interact with a pyrazolopyrimidine derivative and three histone deacetylase (HDAC) inhibitors, valproic acid, trichostatin A, and entinostat (Figure 3). Expectedly, our analysis revealed (6-(4-(2-piperidin-1-ylethoxy)phenyl))-3-pyridin-4-ylpyrazolo(1,5-a)pyrimidine, as a hub chemical, which is structurally similar to pyrazolopyrimidine derivative 10a, showing interactions with 12 resistance related proteins. Interestingly, three HDAC inhibitors have interactions mostly with resistance-associated genes, and are also known as differentiating agents.<sup>17,18</sup> Valproic acid was a hub chemical with the highest number of interactions. Because our previous findings validated proliferative and invasive phenotypes in chemosensitive and resistant cells, respectively, protein chemical network analysis prompted us to use valproic acid to induce differentiation to enhance sensitivity of resistant cells to 10a.

For this purpose, the 10a resistant M221 and M193 cells were treated with various doses of valproic acid (as sodium valproate) alone or in combination with 10a after 24 h of cell seeding. Cell viability was assessed 72 h after treatments. In M221 cells, 20  $\mu$ M 10a treatment resulted in a 55% reduction in viability, whereas little or no decrease in viability was observed with 50, 200, and 800 nM of sodium valproate treatment alone. However, combining 50 nM sodium valproate with 20  $\mu$ M 10a resulted in a 75% decrease in viability. This effect increased to 84% when 200 nM sodium valproate was used. The most significant effect was observed at 20  $\mu$ M 10a

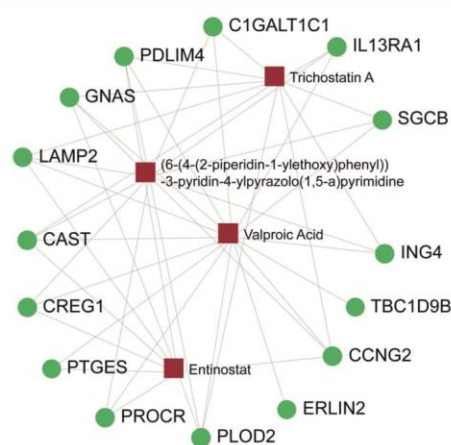


**Figure 3.** Protein–chemical interaction network with chemosensitivity associated genes. A protein–chemical interaction network was constructed using the NetworkAnalyst online tool to identify chemicals that interact with 10a-sensitivity associated proteins. Red squares represent the chemicals; green nodes represent proteins interacting with chemicals, and edges connect proteins to interacting chemicals. Betweenness filter (cutoff = 11.0) was applied to obtain hub chemicals. Valproic acid interacts with 15 proteins represented in the network, while (6-(4-(2-piperidin-1-ylethoxy)phenyl))-3-pyridin-4-ylpyrazolo(1,5-a)pyrimidine interacts with 12 proteins, trichostatin A interacts with 10, and entinostat interacts with 9.

and 0.8 mM sodium valproate treatment, where a 92% decrease in cell viability was seen (Figure 4A). To quantify the degree of synergism, the combination index (CI) was calculated by the Chou–Talalay method in which  $CI < 1$  shows synergistic effects,  $CI = 1$  shows additive effects, and  $CI > 1$  shows antagonistic effects.<sup>19</sup> Combined 10a and sodium valproate treatment showed synergistic effects for all combinations with CI values ranging from 0.766 to 0.603 (Figure 4A). Similarly, the sodium valproate–10a combination showed synergy in the more resistant M193 cells as well (Figure 4B). Twenty micromolar 10a treatment alone, decreased viability to 47% in M193 cells, whereas 0.8, 4, and 6 mM of sodium valproate showed little or no decrease in viability. Adding 0.8 mM sodium valproate to 20  $\mu$ M of 10a resulted in a 75% decrease in viability. Combining 20  $\mu$ M 10a and 4 mM sodium valproate treatment resulted in an 84% reduction in viability. Twenty micromolar 10a and 6 mM sodium valproate combination resulted in 89% decrease in viability. CI values ranging from 0.835 to 0.698 suggested synergism in all combinations (Figure 4B). These results show that sodium valproate acts as a potent sensitizer for 10a in melanoma.

To identify a minimal number of biomarkers that could help classify melanoma subtypes which show differential 10a sensitivity, we selected 4 genes out of the 46, based on highest fold-change ( $\log_2FC > 0.2$  and  $< -0.2$ ) in both directions and significance ( $p$ -value  $< 0.001$ ) between sensitive and resistant cells (highlighted in Figure 1C). TMED7 and PLOD2, whose increased expression correlates with resistance to 10a, along

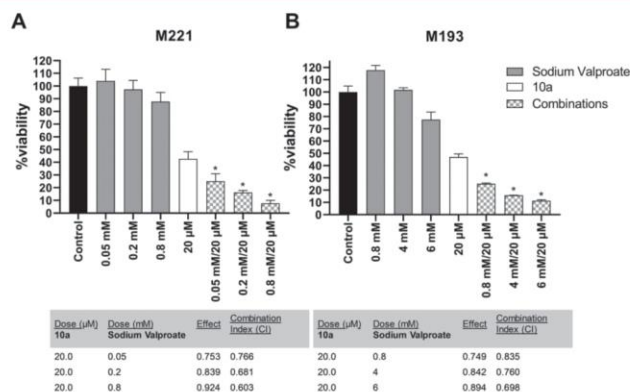


**Figure 4.** Combination treatment of 10a-resistant M221 and M193 melanoma cells with 10a and sodium valproate. (A) After 24 h of cell seeding, M221 cells were treated with different doses (0.05–0.8 mM) of sodium valproate in the absence or presence of 10a. Viable cells were quantified using Cell-Titer Glo assay after 72 h. Combination index values of these combinations suggests synergistic effects in all combinations where CI < 1 indicates synergy. (B) After 24 h of seeding, M193 cells were treated with different doses (0.8–6 mM) of sodium valproate in the absence or presence of 10a. Viable cells were quantified using Cell-Titer Glo assay after 72 h. Combination index values of these combinations suggests synergistic effects in all combinations. \* $p < 0.05$  ( $t$  test).

with XRCC5 and NSUN5 whose increased expression is indicative of sensitivity were chosen to be validated *in vitro*. Expression of these four genes were quantified by RT-qPCR analysis in four sensitive cells (M307, M189, SHM20A2 and M24) and five resistant cells (M34, M433, M2025, M221, and M193). Quantification of gene expression levels of four genes revealed that TMED7 and PLOD2 were upregulated in 10a-resistant cells, whereas XRCC5 and NSUN5 were relatively downregulated, as expected (Figure 5).

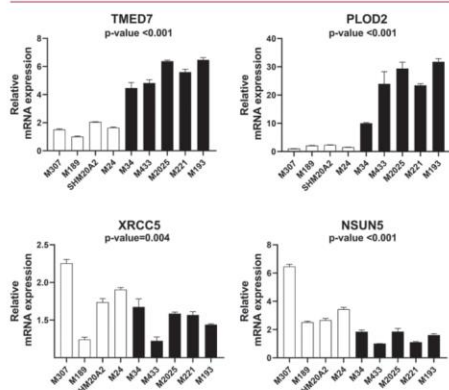


**Figure 5.** Gene expression analysis of candidate biomarkers for 10a chemosensitivity. TMED7, PLOD2, XRCC5, and NSUN5 expressions are quantified by real-time qPCR. GAPDH is used as endogenous control. M307, M189, SHM20A2, and M24 represent sensitive cells and M34, M433, M2025, M221, and M193 are 10a-resistant cells.  $p$ -Values indicate  $t$  test comparison of expression levels between sensitive and resistant cells.

The *in vitro* and *in silico* data we report here indicate that melanoma cells with proliferative or mesenchymal features clearly exhibit distinct sensitivity patterns for the c-Src inhibitor 10a. This is in agreement with our previous work, where a proliferative or invasive phenotype corresponded to 10a-sensitive and resistant cells, respectively.<sup>14</sup> Additionally, combination of 10a with the histone deacetylase inhibitor, valproic acid, indicates that valproic acid can be used to enhance 10a cytotoxicity in these cells. Our expression data show TMED7, PLOD2, XRCC5, and NSUN5 are capable of classifying *in vitro* responses to 10a (Figure S1). Further studies of 10a *in vivo*, combination treatments with valproic acid, and clinical validation of biomarker genes would yield further insights into mechanisms of drug action, degree of cotreatment synergy, and clinical reliability of sensitivity biomarkers.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsmchemlett.9b00679>.

Experimental details for cell culture, bioinformatics analyses, *in vitro* biological assays, combination index analysis, quantification of gene expression levels, hierarchical clustering with *in vitro* data and tables containing *in silico* data (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Ali O. Gure – Department of Molecular Biology and Genetics, Bilkent University, Ankara 06800, Turkey; [orcid.org/0000-0002-4064-8608](https://orcid.org/0000-0002-4064-8608); Email: [agure@bilkent.edu.tr](mailto:agure@bilkent.edu.tr)

### Authors

Baris Kucukkaraduman – Department of Molecular Biology and Genetics, Bilkent University, Ankara 06800, Turkey

**Can Turk** – Faculty of Medicine, Department of Medical Microbiology, Lokman Hekim University, Ankara 06510, Turkey

**Anna L. Fallacara** – Department of Biotechnology, Chemistry and Pharmacy, "Department of Excellence 2018-2022", University of Siena, Siena 53100, Italy

**Murat Isbilen** – DNAFect Genetics Consulting R&D and Biotechnology Inc., Kocaeli 41470, Turkey

**Kerem M. Senses** – Department of Molecular Biology and Genetics, Zonguldak Bulent Ecevit University, Zonguldak 67100, Turkey

**Zeynep O. Ayyildiz** – Department of Genome Sciences and Molecular Biotechnology, Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, Izmir 35340, Turkey

**Muhammad W. Akbar** – Department of Molecular Biology and Genetics, Bilkent University, Ankara 06800, Turkey

**Michal Lotem** – Sharett Institute of Oncology, Hadassah Hebrew University Hospital, Jerusalem 91120, Israel

**Maurizio Botta** – Department of Biotechnology, Chemistry and Pharmacy, "Department of Excellence 2018-2022", University of Siena, Siena 53100, Italy; [orcid.org/0000-0003-0456-6995](https://orcid.org/0000-0003-0456-6995)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsmchemlett.9b00679>

#### Notes

©M.B.: Deceased on August 2, 2019.

The authors declare no competing financial interest.

#### ACKNOWLEDGMENTS

B.K. was funded by the TUBITAK-BIDEB 2211 scholarship program. This work was supported by the Scientific and Technological Research Council of Turkey (112S023) and COST action CM1106.

#### ABBREVIATIONS

MAPK, mitogen-activated protein kinase; TMED7, transmembrane P24 trafficking protein 7; PLOD2, procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2; XRCC5, X-ray repair cross complementing 5; NSUN5, NOP2/Sun RNA methyltransferase 5; PDGFR, platelet-derived growth factor receptor; EPHA, ephrin A; C-MAP, connectivity map; GSEA, gene set enrichment analysis

#### REFERENCES

- (1) Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* **2015**, *161* (7), 1681–96.
- (2) Strojjan, P. Role of radiotherapy in melanoma management. *Radiol. Oncol.* **2010**, *44* (1), 1–12.
- (3) Domingues, B.; Lopes, J. M.; Soares, P.; Populo, H. Melanoma treatment in review. *Immunotargets Ther.* **2018**, *7*, 35–49.
- (4) Irby, R. B.; Yeatman, T. J. Role of Src expression and activation in human cancer. *Oncogene* **2000**, *19* (49), 5636–42.
- (5) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* **2014**, *5*, 3397.
- (6) Lee, J. H.; Pyon, J. K.; Kim, D. W.; Lee, S. H.; Nam, H. S.; Kim, C. H.; Kang, S. G.; Lee, Y. J.; Park, M. Y.; Jeong, D. J.; Cho, M. K. Elevated c-Src and c-Yes expression in malignant skin cancers. *J. Exp. Clin. Cancer Res.* **2010**, *29*, 116.
- (7) Araujo, J.; Logothetis, C. Dasatinib: a potent SRC inhibitor in clinical development for the treatment of solid tumors. *Cancer Treat. Rev.* **2010**, *36* (6), 492–500.

(8) Kluger, H. M.; Dudek, A. Z.; McCann, C.; Ritacco, J.; Southard, N.; Jilaveanu, L. B.; Molinaro, A.; Sznol, M. A phase 2 trial of dasatinib in advanced melanoma. *Cancer* **2011**, *117* (10), 2202–8.

(9) Gangadhar, T. C.; Clark, J. I.; Karrison, T.; Gajewski, T. F. Phase II study of the Src kinase inhibitor saracatinib (AZD0530) in metastatic melanoma. *Invest. New Drugs* **2013**, *31* (3), 769–73.

(10) Schenone, S.; Radi, M.; Musumeci, F.; Brullo, C.; Botta, M. Biologically driven synthesis of pyrazolo[3,4-d]pyrimidines as protein kinase inhibitors: an old scaffold as a new tool for medicinal chemistry and chemical biology studies. *Chem. Rev.* **2014**, *114* (14), 7189–238.

(11) Radi, M.; Brullo, C.; Crespan, E.; Tintori, C.; Musumeci, F.; Biava, M.; Schenone, S.; Dreassi, E.; Zamperini, C.; Maga, G.; Pagano, D.; Angelucci, A.; Bologna, M.; Botta, M. Identification of potent c-Src inhibitors strongly affecting the proliferation of human neuroblastoma cells. *Bioorg. Med. Chem. Lett.* **2011**, *21* (19), 5928–33.

(12) Lotem, M.; Merims, S.; Frank, S.; Hamburger, T.; Nissan, A.; Kadouri, L.; Cohen, J.; Straussman, R.; Eisenberg, G.; Frankenburg, S.; Carmon, E.; Alaiyan, B.; Shneibaum, S.; Ozge Ayyildiz, Z.; Isbilen, M.; Mert Senses, K.; Ron, L.; Steinberg, H.; Smith, Y.; Shiloni, E.; Gure, A. O.; Peretz, T. Adjuvant Autologous Melanoma Vaccine for Macroscopic Stage III Disease: Survival, Biomarkers, and Improved Response to CTLA-4 Blockade. *J. Immunol. Res.* **2016**, *2016*, 8121985.

(13) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15545–50.

(14) Senses, K. M.; Ghasemi, M.; Akbar, M. W.; Isbilen, M.; Fallacara, A. L.; Frankenburg, S.; Schenone, S.; Lotem, M.; Botta, M.; Gure, A. O. Phenotype-based variation as a biomarker of sensitivity to molecularly targeted therapy in melanoma. *MedChemComm* **2017**, *8* (1), 88–95.

(15) Babicki, S.; Arndt, D.; Marcu, A.; Liang, Y.; Grant, J. R.; Maciejewski, A.; Wishart, D. S. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* **2016**, *44* (W1), W147–53.

(16) Zhou, G.; Soufan, O.; Ewald, J.; Hancock, R. E. W.; Basu, N.; Xia, J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **2019**, *47* (W1), W234–W241.

(17) Gottlicher, M.; Minucci, S.; Zhu, P.; Kramer, O. H.; Schimpf, A.; Giavara, S.; Sleeman, J. P.; Lo Coco, F.; Nervi, C.; Pelicci, P. G.; Heinzl, T. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J.* **2001**, *20* (24), 6969–78.

(18) Marks, P. A.; Richon, V. M.; Rifkind, R. A. Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. *J. Natl. Cancer Inst.* **2000**, *92* (15), 1210–6.

(19) Chou, T. C.; Talalay, P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv. Enzyme Regul.* **1984**, *22*, 27–55.

## Plasticity as an Underlying Mechanism of Tumor Heterogeneity in Breast Cancer

Authors: Muhammad Waqas Akbar<sup>1</sup>, Murat Isbilen<sup>2</sup>, Baris Kucukkaraduman<sup>1</sup>, Secil Demirkol Canli<sup>3</sup>, Ege Dedeoglu<sup>1</sup>, Shila Azizolli<sup>1</sup>, Isli Cela<sup>1</sup>, Abbas Guven Akcay<sup>1</sup>, Hasim Hakanoglu<sup>1</sup>, Ali Osmay Gure<sup>1,\*</sup>

<sup>1</sup> Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey

<sup>2</sup> DNAFect Genetics Consulting R&D and Biotechnology Inc., Kocaeli, Turkey

<sup>3</sup> Molecular Pathology Application and Research Center, Hacettepe University, Ankara, Turkey

\* agure@bilkent.edu.tr

### Abstract

Breast cancer shows plasticity in terms of classification. Upon drug treatment and metastasis some tumors switch to another subtype leading to loss of response to therapy. In this study, we ask the question which molecular subclasses of breast cancer are more switchable upon drug therapy and metastasis. We used in silico data to classify breast cancer tumors in PAM50 molecular classes before treatment and after treatment using gene expression data. Similar analysis was performed for primary tumors and their metastatic growth. Our analysis showed that in both scenarios some breast tumors shift from one class to another. This suggests that patients who underwent chemotherapy but resulted in relapse or/and metastasis should be retyped for molecular subclass so that treatment protocol should be adopted according to those subtypes. Additionally, 20 genes were identified as biomarkers for metastasis in breast cancer.

### Introduction

Breast cancer (BC) is the second leading cause for cancer related mortality in women after lung cancer<sup>1,2</sup>. In clinics, breast cancer is classified using estrogen receptor (ER), progesterone receptor (PGR) and ErbB2 Receptor Tyrosine Kinase 2 (ERBB2, previously known as HER2) statuses through immunohistochemistry (IHC classes). Moreover, BC is divided in Luminal A, Luminal B, HER2 enriched, Basal and Normal like subgroups by molecular classification<sup>3</sup>. Metastatic breast cancer (MBC) is the type of BC in which cancer cells leave primary cancer tissue and spread to other parts of the body. MBC remains an incurable disease as its median overall survival is about 3 years and 25% patients show only 5 year survival<sup>4</sup>.

MBC consists of a series of sequential steps, including transfer of cells from primary tumor to circulation (intravasation), survival in circulation, extravasation in a new organ, colonization and metastatic growth. Metastatic cells go through epithelial to mesenchymal transition (EMT) to achieve metastasis<sup>5</sup>. This metastatic cascade results in switching the IHC status of BC in some samples, assigning metastatic growth a different class for which primary therapy might not be suitable and a different protocol should be opted to combat this growth<sup>6</sup>. Previously similar subtype switching is observed in PAM50 classes as well<sup>7</sup>. In this paper we address this switching as plasticity. Additionally, this subtype plasticity is also seen among pre- and post-treated tumors due to epithelial to mesenchymal transition<sup>8</sup>.

In this paper we used public gene expression datasets to determine if there is a subtype switch between pre- and post- therapy patient tumor samples and also if this switch is present between primary and metastatic tissue. Moreover MBC biomarkers were identified using primary tumor expression data which can be used to determine if a breast cancer is capable of metastasis or not.

## Methods

### PAM50 plasticity

Datasets GSE10281, GSE32518, GSE32072 (patient samples before and after treatment), GSE32518 (patient samples collected with fine needle aspiration and core needle biopsy), GSE125989, GSE110590 (patient samples for primary and metastatic tissues) were downloaded from GEO NCBI database. All datasets were “justRMA” normalized using BRB Array Tools<sup>9</sup>. PAM50 signature was used to classify breast cancer into 5 molecular subgroups: Luminal A (LumA), Luminal B (LumB), HER2 enriched (HER2), Basal and Normal like (Normal) as described previously<sup>10</sup> for each dataset. Paired samples were matched for pre- and post- therapy, with fine needle aspiration and core needle biopsy, or primary and metastatic tissue to compare PAM50 plasticity in these samples.

### Metastasis biomarkers

For metastasis-free survival analysis, GSE2603 (n=82, treatment not disclosed), GSE6532 ((n= 293, platform: Affymetrix Human Genome U133A and B Array, Tamoxifen treated), (n=87, Affymetrix Human Genome U133 Plus 2.0 Array)), GSE7390 (n=198, systemically untreated patients), GSE11121 (n=200, only surgery was performed), GSE12276 (n=204, adjuvant chemotherapy treated), GSE20685 (n=327, adjuvant chemotherapy treated) and GSE58812 (n=108, adjuvant chemotherapy treated) were downloaded. All datasets were “justRMA” normalized using BRB Array Tools. We performed cox proportional hazard analysis using “Survival” package built in R software for all datasets<sup>11,12</sup>. Genes that showed same direction in all datasets and were significant for 5/8 analysis were selected as metastasis biomarkers. 40 probesets were found significant and to select top 10 good and bad prognostic genes, genes were ranked by cox p value in each dataset individually for both directions and these ranks were summed. Top 10 genes were then selected from good prognostic genes based upon the least ranksum score and another 10 genes were selected from bad prognostic genes based on the same criteria. For genes which were significant with 2 or more probesets, only one probeset with the least ranksum score was selected.

## Results

### PAM50 plasticity

PAM50 is also prone to show different classes, as analyzed in dataset GSE32518, where authors collected fine needle aspiration and core needle biopsies from matching patients (N=37). When these samples were typed for PAM50, 78% of these samples (n=29) showed same phenotype in both biopsy methods while 22% samples (n=8) showed discordant subtypes. One HER2 sample by core needle biopsy was identified as Basal and another one was identified as LumB. One sample was identified as LumA (Core needle biopsy) and LumB (Fine needle aspiration). Most discordant results were for Normal subtype (Core needle biopsy) where 4 samples were identified Basal and 1 sample as LumA (Table 1). This PAM50 discordance can be attributed to tumor heterogeneity.

Plasticity exists between PAM50 classes, and tumor subtype can switch to another upon drug treatment through EMT and immunological status. To study PAM50 plasticity, we analyzed different datasets in which matched tumor samples gene expression data was available for both pre- and post-chemotherapy. When Letrozole pre- and post-treated samples (GSE10281) were compared for PAM50 class plasticity, 39% samples (7/18) retained their PAM50 class before and after Letrozole therapy and 61% (11/18) samples changed their class. Basal (n=2) and LumB (n=1) subtype was most stable in this dataset. One HER2, 2 LumA and 1 normal subtype samples retained their subtype while 3 LumA, 3 HER2 and 4

normal pre therapy samples switched to another subtype. Most of these samples (n=6) switched to LumB and 4 samples switched to Basal subtype (Table 2). Authors reported in their paper that upon chemotherapy treatment, residual tumors attain mesenchymal properties as expected<sup>13</sup>. Among patients treated with Anthracyclines and Taxanes based neoadjuvant chemotherapy (GSE32518), 2 basal samples, 1 HER2, 3 LumA, 3 lumB and 1 Normal retained their subtype and 2 basal, 3 HER2, 4 LumA, 8 LumB and 1 normal switched their phenotype (Table 3). Most samples switched their subtype to normal like tumors. Authors reported that genes related to extracellular metabolism, angiogenesis and developmental processes are related to chemoresistance in these patients<sup>14</sup>. In another dataset (GSE32072), where gene expression data for 21 cancer samples was available for pre- and post-neoadjuvant therapy (9 patients anthracycline based therapy, 8 patients anthracycline/taxane based therapy and 2 patients trastuzumab therapy). Additionally, 4 patients were also included which were not treated with any neoadjuvant therapy as control. Upon PAM50 subtyping, patients who did not receive any neoadjuvant chemotherapy, retained their classes (2 basal and 2 LumA). Interestingly, all basal (n=5), Her2 (n=4) and LumB (1) retained their subtype. Out of 6 pre-therapy lumA samples only one changed to LumB subtype. Normal subtype samples changed the most to other subtypes in this dataset (N=4) and only one sample retained its phenotype (Table 4). Authors reported enrichment of energy and metabolism related genes in residual samples while immune related genes were depleted<sup>15</sup>.

Breast cancer metastasis can be molecularly distinct from their primary tissues<sup>16</sup>. Metastatic samples can switch to another PAM50 class when compared with primary tissue class. In dataset GSE125989, 16 matched tissues were available for both brain metastasis and primary site. These samples were PAM50 subtyped. 50% (8) of them retained their phenotype when primary site vs metastasis PAM50 classes were compared. One Basal changed to Her2 and 1 Her2 changed to basal. Major subtype switching was observed in Normal primary tissue phenotype, where 3 metastasized tissues became basal and 2 became HER2 (Table 5). Authors showed that when gene expression of metastasized tissue was compared with primary tissue, mesenchymal markers were up regulated while epithelial markers were downregulated. Additionally tumor infiltrating lymphocytes, B cells and dendritic cell gene signatures were downregulated in metastatic tissues. Dataset GSE110590 contains gene expression for 16 primary breast cancer tissues and their 67 matched metastasis. 55% (37/67) metastatic tissues retained the same PAM50 class when compared with primary site. Most stable classes were basal where 85% (23/27) metastatic sites retained basal phenotype and HER2 100% (4/4). While LumA showed 32% (10/31) and Normal showed 0% (0/5) consistency. Most primary samples when metastasized switched to Basal and LumB class (Table 6).

In another study, authors compared 123 patients' primary tissues with metastatic tissues for PAM50 switching. Metastatic tissues included skin, lymph node, liver, bone, lung, peritoneum and others. Subtype concordance for Basal was highest (100%), followed by HER2 (76.92%) and LumB (70%). The most switchable subtype was LumA where 55.32% sample switched to luminal B and HER2<sup>7</sup>.

Above analysis suggests PAM50 subtype switch between different classes both due to chemotherapy treatment and metastasis. Therefore subtyping for residual cancer or metastasis should be re-evaluated. Additionally, basal phenotype is the most stable phenotype and LumA is the most unstable one. Moreover PAM50 classification is subjected to sampling method as we show that different sampling techniques can result in different classification of the same samples.

### **Biomarkers for metastasis**

To identify biomarkers for metastasis, we used 7 cohorts for which metastasis-free survival data was reported. Treatment protocol and patient numbers for each cohort are described in methods section. We

selected 10 good and 10 bad prognostic genes. All selected 20 genes and their respective probeset IDs are shown in figure 1.

Our selected genes can predict metastasis-free survival or distant metastasis-free survival in several cohorts. Next we used these genes to cluster patients in metastatic prone or resistant groups with the aim to identify enriched gene sets among these using GSEA<sup>17</sup>. Surprisingly, no common gene sets enriched in metastatic prone and resistant groups were found, suggesting that different mechanisms like EMT, inflammation, cell cycle and immune response are involved in metastasis (data not shown).

### Good prognostic genes

Among the 20 genes selected, 10 were good prognostic. Epoxide hydrolase 2 (EPHX2) is largely involved in metabolism of epoxy fatty acids and plays an important part in cardiovascular biology, renal function, inflammation response and pain<sup>18</sup>. Previously, EPHX2 has been associated with metastasis suppression in breast cancer and its expression was four-fold higher in lymph node negative tumors as compared with lymph node positive tumors<sup>19</sup>. In another meta-analysis for breast cancer metastasis, EPHX2 was identified as metastasis suppressor gene and its biological functions were associated with drug metabolism, inflammatory response and reactive oxygen species metabolism<sup>20</sup>. So, it might be possible that EPHX2's diverse role in metabolism of carcinogens might play a role in metastasis of breast cancer. Previously high glucose levels were associated with increased breast cancer proliferation and metastasis<sup>21</sup> and interestingly specificity protein 1 (sp1) is shown to downregulate EPHX2 under increased glucose levels<sup>22</sup>. EPHX2 role has also been associated with recurrence in breast cancer<sup>23</sup>.

Thymocyte nuclear protein 1 (THYN1), also known as THY28, is a highly conserved protein from bacteria to mammals and plays an important role in B cell development<sup>24</sup> and its downregulation has been associated with induction of apoptosis in Ramos B lymphoma cells<sup>25</sup>. However its role in breast cancer is not well studied. Previously, another study on breast cancer metastasis has associated its role as good prognosis marker as they used the same cohort GSE6532 (also used in this study). In another study, THYN1 role was associated with recurrence in breast cancer<sup>23</sup>. THYN1 is also negatively correlated with numerical chromosomal heterogeneity in NCI60 cancer cell lines. In this study, authors showed that chromosomal instability is associated with EMT, cancer invasiveness and metastasis<sup>26</sup>.

ADP ribosylation factor like GTPase 6 interacting protein 5 (ARL6IP5), also known as JWA, has been shown previously as microtubule associated protein and has been related to differentiation, apoptosis and response to stress stimulators and its expression is mediated by ER<sup>27</sup>. In NIH3T3 and HELF cell lines, increased H<sub>2</sub>O<sub>2</sub> levels resulted in increased expression of ARL6IP5 expression indicating its role in oxidative stress<sup>28</sup>. In breast cancer, knockdown of ARL6IP5, resulted in increased invasion and migration along with decreased apoptosis in MDA-MB-231 cells<sup>29</sup>. In another study, authors showed that expression of CXCR4 (breast cancer metastatic marker) was inversely proportional to ARL6IP5 and ARL6IP5 decreases breast cancer migration and invasion by proteomic degradation of CXCR4<sup>30</sup>. Moreover, ARL6IP5 has been shown to suppress metastasis of melanoma cells as well as cancer cell adhesion, invasion and metastasis in gastric cells, similar to our analysis<sup>31-33</sup>.

CAMP responsive element binding protein like 2 (CREBL2) is identified as a metabolic regulator<sup>34</sup>. CREBL2 was shown to be expressed by T47D cell line when compared with MDA-MB-435 (a highly metastatic cell line)<sup>35</sup>. CREBL2 expression is decreased in inflammatory breast cancer (an aggressive subtype of breast cancer)<sup>36</sup>. CREBL2 was found to be downregulated in OVCAR paclitaxel resistant cell line (ovarian cancer cell line) when compared with controls<sup>37</sup>. CREBL2 was also found downregulated among grade 3 breast cancer samples<sup>38</sup>. In another study, to identify genes differentially expressed

between patients who suffered from distant metastasis and ones who did not, CREBL2 was downregulated in metastatic group<sup>39</sup>.

Purinergic receptor P2Y13 (P2RY13) is a G-protein-coupled receptor and it is downregulated upon EGF- and hypoxia- induced EMT in breast cancer cells<sup>40,41</sup>. Additionally, overexpression of P2RY13 is associated with better recurrence free survival in hepatocellular carcinoma<sup>42</sup>.

Alpha-L-Fucosidase 1 (FUCA1) is a lysosomal enzyme involved in cleavage of terminal fructose residue. Its lower expression has been associated with aggressive tumors such as breast, neuroblastoma and colorectal cancer. Moreover its reduced expression is related to shorter recurrence and cancer free survival in lymph node positive breast cancer luminal B patients<sup>43</sup>. Its lower expression is also related to metastasis in solid tumors<sup>44</sup>. One study showed that FUCA1 functions downstream of p53 and can contribute to EGFR signaling repression by removing fucose from EGFR and thus is a marker of lower survival in breast cancer<sup>45</sup>. Among triple negative breast cancer patients, FUCA1 downregulation is associated with decreased survival through modulation of glycosylation status<sup>46</sup>.

Kinesin family member 13B (KIF13B) belongs to Kinesin 3 family. KIF13B is downregulated upon estrogen stimulation of breast cancer cell proliferation<sup>47</sup>. And its expression is upregulated in paclitaxel resistant MCF7 cells when compared with control cells<sup>48</sup>. Although KIF13B is not studied well in cancer, its role was emphasized as VEGFR2 trafficker in endothelial cells previously. Thus it plays an important role in angiogenesis and its depletion leads to reduced VEGF induced capillary tube formation, endothelial cells migration and neovascularization in mice<sup>49</sup>.

Progesterone receptor (PGR) status has been described in previous sections as it goes through plasticity during metastasis. In one study, authors found 46% tumors shifting from PR positive status to negative status in metastatic tissue, although in the same study they also saw the opposite trend for 16% tumors as well<sup>50</sup>. In another study, 40.4% discordance was found between primary and metastatic tissue regarding PGR status and 74% discordance was due to loss of PGR status<sup>51</sup>.

DIX domain containing 1 (DIXDC1) gene is a regulator of Wnt signaling pathway and its expression has been associated with good prognosis in breast cancer as well as metastasis repression in lung adenocarcinoma as expected<sup>52,53</sup>. Knockdown of DIXDC1 in glioma cells has been shown to inhibit proliferation and migration of these cells<sup>54</sup>. Interestingly lower expression of DIXDC1 has been associated with worse outcomes in hepatocellular carcinoma<sup>55</sup> and intestinal type gastric carcinoma<sup>56</sup>.

Joining chain of multimeric IgA and IgM (JCHAIN), previous symbol IGJ, is downregulated in ductal carcinoma *in situ* and invasive ductal carcinoma when compared with normal stroma<sup>57</sup>. Additionally, JCHAIN was previously shown to be downregulated in both tumor stroma and tumor epithelium in inflammatory breast cancer (aggressive form of breast cancer related to metastasis) when compared with non-inflammatory breast cancer<sup>36</sup>.

#### **Bad prognostic genes**

Serine dehydratase (SDS) converts L-serine to pyruvate and ammonia. It has not been studied extensively in breast cancer. In one study, its expression was shown to be higher in tumor samples when compared with normal samples<sup>58</sup>.

N-Alpha-Acetyltransferase 10, NatA Catalytic Subunit (NAA10), also known as ARD1, is related to post-translational modifications and has been previously shown to increase breast cancer proliferation upon overexpression<sup>59</sup>. Its expression is increased in breast tumor cells when compared with normal adjacent

cells<sup>60</sup>. In one study, authors related its expression with lymph node metastasis in breast cancer<sup>60</sup>. Surprisingly in another study, it was reported as the inhibitor of metastasis by inhibiting cell motility<sup>61</sup>.

Centromere protein E (CENPE), with chromosome misalignment being the hallmark of cancer, aids the proliferation of tumor cells via its involvement in capturing and positioning the chromosomes to mitotic spindle<sup>62,63</sup>. The inhibition of CENP-E causes the death and remediation of triple negative breast cancer cells<sup>64,65</sup>.

Argonaute-2 (AGO2) plays an indispensable role in the post-transcriptional downregulation of gene expression through guided small-RNAs. It is one of the four members of the Argonaute protein family which is highly conserved amongst different species and is the only member of the family to have catalytic activity and therefore, is the most important for gene silencing along with the RISC complex<sup>66,67</sup>. AGO2 has been linked to different types of cancers previously<sup>68,69</sup>. AGO2 expression was shown to increase in Her2 and basal subtype. Additionally increase in AGO2 expression showed reduced relapse free survival. In the same research it was indicated that cancer cell-lines showed weak AGO2 expression as opposed to the tissue samples, suggesting a role of the tumor microenvironment in AGO2 expression<sup>70</sup>. AGO2 has also been shown to be highly expressed in a subset of ERα+ breast tumors and is correlated with poor prognosis in patients. It was shown that the endocrine resistance of ER+/PGR- tumors were highly correlated with AGO2 overexpression<sup>71</sup>. Cells transfected with AGO2 were shown to have a more aggressive phenotype with higher proliferative capabilities and higher expression of pro-oncogenic microRNAs<sup>72</sup>.

E2F8 is the eighth member of the E2F transcription factors, which are expressed in a variety of tissues and is involved in regulatory functions of cellular proliferation, differentiation, cell apoptosis, DNA repair & cell cycle<sup>73,74</sup>. E2F8 plays a major role in cell-cycle progression, especially to the G1 to S phase transition by binding to E2F1, repressing its expression and therefore inhibiting its effects<sup>75</sup>. This property has been reported to affect breast cancer tumorigenesis as well. It accomplishes this by transcriptionally activating CCNE1 and CCNE2 leading to G1/S phase transition resulting in increased proliferation<sup>76</sup>. E2F8 has also been shown to be overexpressed in patient data obtained from TCGA. Patients that had higher expression of E2F8 showed significantly poorer prognosis<sup>77</sup>. Additionally it has been recently shown that post-transcriptional regulation of E2F8 by RNA binding proteins in breast cancer cell lines, increases the proliferative capabilities of breast cancer<sup>78</sup>. It is clear that E2F8 overexpression increases the invasive properties of breast cancer and results in poorer prognosis of breast cancer patients when overexpressed.

BUB1 Mitotic Checkpoint Serine/Threonine Kinase B (BUB1B) has a significant relation with both stage and grade of the tumor. Overexpression of BUB1B is more often noticed in high grade breast cancer and it is correlated with shorter distant metastasis-free survival<sup>79</sup>. BUB1B's expression values have been shown to solely divide the breast cancer patients into metastatic and non-metastatic<sup>80</sup>.

Transmembrane Protein 208 (TMEM208) codes for a transmembrane protein located on endoplasmic reticulum. TMEM208 overexpression has been shown to inhibit ER stress and autophagy<sup>81</sup>. Also, TMEM208 transcription is affected by oxygen levels, especially by hypoxic conditions under which HIF-1α binds to its promoter, resulting in an increase in transcript levels<sup>82</sup>. Hypoxic conditions have been shown to induce invasiveness of breast cancer cells under the orchestration of HIF-1α<sup>83</sup>.

Proteasome 26S Subunit, Non-ATPase 2 (PSMD2) codes for a protein part of the proteasome subunit of the ubiquitin proteasome system (UPS). This gene is upregulated in breast cancer cells as compared with normal adjacent tissue by more than 3 fold<sup>84</sup>. High levels of PSMD2 correlate with shorter overall survival and distant metastasis free survival and the gene is also part of a metastatic signature that relates

to poorer prognosis<sup>85,86</sup>. Silencing of PSMD2 inhibits breast cancer cell proliferation, tumor growth and increases apoptosis<sup>85</sup>.

Tubulin Alpha 1b (TUBA1B) is a member of tubulin family proteins, which are major components of microtubules. TUBA1B is involved in a diverse spectrum of biological activities including cell adhesion, movement, replication and division. Although its role is undetermined in breast cancer and metastatic progression, it was shown to be associated with poor survival in liver cancer<sup>87</sup>.

Ubiquitin Conjugating Enzyme E2 A (UBE2A), also known as RAD6A, is a ubiquitin conjugating enzyme which is involved in protein metabolism. Studies showed that UBE2A is upregulated in aggressive breast cancers and its expression correlates with metastasis<sup>88 89</sup>. In one study, authors also revealed that increased RAD6 levels are associated with cancer stem cell phenotype in ovarian cancer resulting in disease recurrence and metastasis<sup>90</sup>.

### Conclusion

Our results suggest that PAM50 plasticity exists in response to treatment. Moreover, metastasized tissue shows different PAM50 class when compared with primary tissue. So, there is a need to reassess the classes in case of relapse and metastasis. For example patients with HER2 enriched phenotype are treated with targeted therapy, but if these tumors switch to a luminal phenotype upon relapse or metastasis, then a second line of treatment can be added to deal with this switch<sup>7</sup>.

Additionally, our suggested metastatic biomarkers can be used to better predict metastatic events in breast cancer patients. Previously in literature several studies have identified metastatic biomarkers in breast cancer<sup>91-93</sup>. But here our markers show prognostic importance in 7 different cohorts.

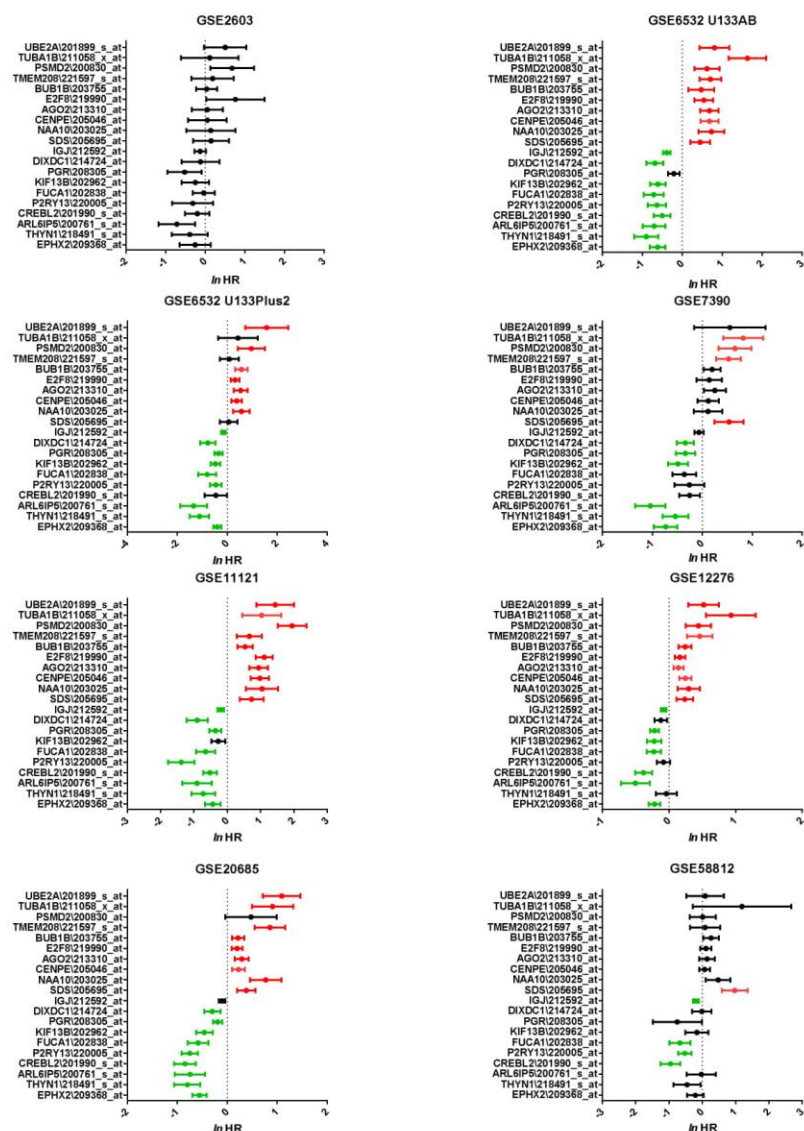


Figure 1: Metastasis-free survival analysis: 20 genes (10 good prognostic, 10 bad prognostic) were selected as metastasis biomarkers. Gene names and probe IDs are shown on y axis. Y axis shows Hazard ratio (HR) in natural logarithm with 95% CI. Genes which are good prognostic and significant are shown in as green while bad prognostic and significant genes are shown in red color. If a gene was not significant it is shown in black.

Table 1: Comparison of PAM50 subtyping for samples collected through fine needle aspiration and core needle biopsy (GSE32518).

|                      |        | Fine needle aspiration |      |      |      |        |
|----------------------|--------|------------------------|------|------|------|--------|
|                      |        | Basal                  | HER2 | LumA | LumB | Normal |
| Core needle biopsies | Basal  | 7                      | 0    | 0    | 0    | 0      |
|                      | HER2   | 1                      | 5    | 0    | 1    | 0      |
|                      | LumA   | 0                      | 0    | 9    | 1    | 0      |
|                      | LumB   | 0                      | 0    | 0    | 7    | 0      |
|                      | Normal | 4                      | 0    | 1    | 0    | 1      |

Table 2: Comparison of PAM50 subtypes pre- and post-Letrozole therapy (GSE10281).

|             |        | Post-therapy |      |      |      |        |
|-------------|--------|--------------|------|------|------|--------|
|             |        | Basal        | HER2 | LumA | LumB | Normal |
| Pre-therapy | Basal  | 2            | 0    | 0    | 0    | 0      |
|             | HER2   | 1            | 1    | 0    | 2    | 0      |
|             | LumA   | 0            | 0    | 2    | 3    | 0      |
|             | LumB   | 0            | 0    | 0    | 1    | 0      |
|             | Normal | 3            | 0    | 1    | 1    | 1      |

Table 3: Comparison of PAM50 subtypes pre- and post-Anthracyclines and Taxanes based neoadjuvant chemotherapy (GSE32518).

|             |        | Post-therapy |      |      |      |        |
|-------------|--------|--------------|------|------|------|--------|
|             |        | Basal        | HER2 | LumA | LumB | Normal |
| Pre-therapy | Basal  | 2            | 0    | 0    | 0    | 2      |
|             | HER2   | 1            | 1    | 0    | 1    | 1      |
|             | LumA   | 0            | 0    | 3    | 0    | 4      |
|             | LumB   | 1            | 0    | 4    | 3    | 3      |
|             | Normal | 0            | 0    | 1    | 0    | 1      |

Table 4: Comparison of PAM50 subtypes pre- and post- neoadjuvant chemotherapy (GSE32072).

|             |        | Post-therapy |      |      |      |        |
|-------------|--------|--------------|------|------|------|--------|
|             |        | Basal        | HER2 | LumA | LumB | Normal |
| Pre-therapy | Basal  | 5            | 0    | 0    | 0    | 0      |
|             | HER2   | 0            | 4    | 0    | 0    | 0      |
|             | LumA   | 0            | 0    | 5    | 1    | 0      |
|             | LumB   | 0            | 0    | 0    | 1    | 0      |
|             | Normal | 2            | 1    | 1    | 0    | 1      |

Table 5: Comparison of PAM50 subtyping for primary tissues vs. brain metastatic tissues (GSE125989).

|                |        | Metastasized tissue |      |      |      |        |
|----------------|--------|---------------------|------|------|------|--------|
|                |        | Basal               | HER2 | LumA | LumB | Normal |
| Primary tissue | Basal  | 2                   | 1    | 0    | 0    | 0      |
|                | HER2   | 1                   | 2    | 0    | 0    | 1      |
|                | LumA   | 0                   | 0    | 4    | 0    | 0      |
|                | LumB   | 0                   | 0    | 0    | 0    | 0      |
|                | Normal | 3                   | 2    | 0    | 0    | 0      |

Table 6: Comparison of PAM50 subtyping for primary tissues vs. metastatic tissues (GSE110590).

| Sample name | Primary tissue | PAM50 metastasis |      |      |      |        |
|-------------|----------------|------------------|------|------|------|--------|
|             |                | Basal            | HER2 | LumA | LumB | Normal |
| A11         | Basal          | 5                | 0    | 0    | 0    | 0      |
| A15         | Basal          | 2                | 0    | 0    | 0    | 3      |
| A1          | Basal          | 4                | 0    | 0    | 0    | 1      |
| A23         | Basal          | 5                | 0    | 0    | 0    | 0      |
| A5          | Basal          | 2                | 0    | 0    | 0    | 0      |
| A7          | Basal          | 5                | 0    | 0    | 0    | 0      |
| A8          | HER2           | 0                | 4    | 0    | 0    | 0      |
| A12         | LumA           | 0                | 0    | 2    | 4    | 0      |
| A17         | LumA           | 2                | 0    | 0    | 0    | 0      |
| A26         | LumA           | 3                | 0    | 0    | 0    | 0      |
| A28         | LumA           | 0                | 0    | 3    | 3    | 0      |
| A2          | LumA           | 0                | 1    | 0    | 2    | 0      |
| A30         | LumA           | 5                | 0    | 0    | 0    | 0      |
| A34         | LumA           | 0                | 0    | 4    | 0    | 0      |
| A4          | LumA           | 0                | 1    | 1    | 0    | 0      |
| A20         | Normal         | 5                | 0    | 0    | 0    | 0      |

## References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018;68:394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA: a cancer journal for clinicians* 2019;69:7-34.
3. Harbeck N, Penault-Llorca F, Cortes J, et al. Breast cancer. *Nature Reviews Disease Primers* 2019;5:66.
4. Cardoso F, Senkus E, Costa A, et al. 4th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 4). *Annals of Oncology* 2018;29:1634-1657.
5. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* 2017;168:670-691.
6. Sari E, Guler G, Hayran M, et al. Comparative study of the immunohistochemical detection of hormone receptor status and HER-2 expression in primary and paired recurrent/metastatic lesions of patients with breast cancer. *Medical Oncology* 2011;28:57-63.
7. Cejalvo JM, de Dueñas EM, Galván P, et al. Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer. *Cancer research* 2017;77:2213-2221.
8. Akbar MW, Isbilen M, Belder N, et al. A Stemness and EMT Based Gene Expression Signature Identifies Phenotypic Plasticity and is A Predictive but Not Prognostic Biomarker for Breast Cancer. *Journal of Cancer* 2020;11:949.
9. Simon R, Lam A, Li M-C, et al. Analysis of gene expression data using BRB-array tools. *Cancer informatics* 2007;3:117693510700300022.
10. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology* 2009;27:1160.
11. Team RC. R Foundation for Statistical Computing; Vienna, Austria: 2015. *R: A language and environment for statistical computing* 2018:2013.
12. Therneau T, Grambsch P. Modeling survival data: Extending the Cox model (pp. 69–75): New York, NY: Springer Science & Business Media, 2013.
13. Creighton CJ, Li X, Landis M, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proceedings of the National Academy of Sciences* 2009;106:13820-13825.
14. Vera-Ramirez L, Sanchez-Rovira P, Ramirez-Tortosa CL, et al. Transcriptional shift identifies a set of genes driving breast cancer chemoresistance. *PLoS One* 2013;8.
15. Gonzalez-Angulo AM, Iwamoto T, Liu S, et al. Gene expression, molecular class changes, and pathway analysis after neoadjuvant systemic therapy for breast cancer. *Clinical Cancer Research* 2012;18:1109-1119.
16. Vecchi M, Confalonieri S, Nuciforo P, et al. Breast cancer metastases are molecularly distinct from their primary tumors. *Oncogene* 2008;27:2148-2158.
17. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005;102:15545-15550.
18. Tanaka H, Kamita SG, Wolf NM, et al. Transcriptional regulation of the human soluble epoxide hydrolase gene EPHX2. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 2008;1779:17-27.
19. Ellsworth R, Valente A, Field L, et al. Abstract P4-06-09: Genetic Signature Discriminating Metastatic from Non-Metastatic Small Tumors: AACR, 2010.

20. Thomassen M, Tan Q, Kruse TA. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast cancer research and treatment* 2009;113:239-249.
21. Sun S, Sun Y, Rong X, et al. High glucose promotes breast cancer proliferation and metastasis by impairing angiotensinogen expression. *Bioscience reports* 2019;39.
22. Oguro A, Oida S, Imaoka S. Down-regulation of EPHX2 gene transcription by Sp1 under high-glucose conditions. *Biochemical Journal* 2015;470:281-291.
23. Naoi Y, Kishi K, Tanei T, et al. Development of 95-gene classifier as a powerful predictor of recurrences in node-negative and ER-positive breast cancer patients. *Breast cancer research and treatment* 2011;128:633-641.
24. Fujita T, Kitaura F, Fujii H. A critical role of the Thy28-MYH9 axis in B cell-specific expression of the Pax5 gene in chicken B cells. *PLoS One* 2015;10.
25. Jiang X, Toyota H, Yoshimoto T, et al. Anti-IgM-induced down-regulation of nuclear Thy28 protein expression in Ramos B lymphoma cells. *Apoptosis* 2003;8:509-519.
26. Roschke AV, Glebov OK, Lababidi S, et al. Chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin maintenance. *Neoplasia (New York, NY)* 2008;10:1222.
27. Chen R, Li A, Zhu T, et al. JWA—A novel environmental-responsive gene, involved in estrogen receptor-associated signal pathway in MCF-7 and MDA-MB-231 breast carcinoma cells. *Journal of Toxicology and Environmental Health, Part A* 2005;68:445-456.
28. Chen R, Qiu W, Liu Z, et al. Identification of JWA as a novel functional gene responsive to environmental oxidative stress induced by benzo [a] pyrene and hydrogen peroxide. *Free Radical Biology and Medicine* 2007;42:1704-1714.
29. Chen X, Feng J, Ge Z, et al. Effects of the JWA gene in the regulation of human breast cancer cells. *Molecular medicine reports* 2015;11:3848-3853.
30. Xu L, Cheng L, Yang F, et al. JWA suppresses the invasion of human breast carcinoma cells by downregulating the expression of CXCR4. *Molecular medicine reports* 2018;17:8137-8144.
31. Lu J, Tang Y, Cheng Y, et al. ING4 regulates JWA in angiogenesis and their prognostic value in melanoma patients. *British journal of cancer* 2013;109:2842-2852.
32. Chen Y, Xia X, Wang S, et al. High FAK combined with low JWA expression: clinical prognostic and predictive role for adjuvant fluorouracil-leucovorin-oxaliplatin treatment in resectable gastric cancer patients. *Journal of gastroenterology* 2013;48:1034-1044.
33. Bai J, Zhang J, Wu J, et al. JWA regulates melanoma metastasis by integrin  $\alpha$  V  $\beta$  3 signaling. *Oncogene* 2010;29:1227-1237.
34. Tiebe M, Lutz M, Tiebe DS, et al. Creb12 regulates cell metabolism in muscle and liver cells. *Scientific Reports* 2019;9:1-12.
35. Mellick AS, Day CJ, Weinstein SR, et al. Differential gene expression in breast cancer cell lines and stroma—tumor differences in microdissected breast cancer biopsies revealed by display array analysis. *International journal of cancer* 2002;100:172-180.
36. Boersma BJ, Reimers M, Yi M, et al. A stromal gene signature associated with inflammatory breast cancer. *International journal of cancer* 2008;122:1324-1332.
37. Duan Z, Lamendola DE, Duan Y, et al. Description of paclitaxel resistance-associated genes in ovarian and breast cancer cell lines. *Cancer chemotherapy and pharmacology* 2005;55:277-285.
38. Cava C, Bertoli G, Ripamonti M, et al. Integration of mRNA expression profile, copy number alterations, and microRNA expression levels in breast cancer to improve grade definition. *PLoS one* 2014;9.

39. Myhre S, Mohammed H, Tramm T, et al. In silico ascription of gene expression differences to tumor and stromal cells in a model to study impact on breast cancer outcome. *PLoS One* 2010;5.
40. Davis FM, Kenny PA, Soo ET, et al. Remodeling of purinergic receptor-mediated Ca<sup>2+</sup> signaling as a consequence of EGF-induced epithelial-mesenchymal transition in breast cancer cells. *PLoS one* 2011;6.
41. Azimi I, Beilby H, Davis FM, et al. Altered purinergic receptor-Ca<sup>2+</sup> signaling associated with hypoxia-induced epithelial-mesenchymal transition in breast cancer cells. *Molecular oncology* 2016;10:166-178.
42. Maynard JP, Lee J-S, Sohn BH, et al. P2X3 purinergic receptor overexpression is associated with poor recurrence-free survival in hepatocellular carcinoma patients. *Oncotarget* 2015;6:41162.
43. Bonin S, Parascandolo A, Aversa C, et al. Reduced expression of  $\alpha$ -L-Fucosidase-1 (FUCA-1) predicts recurrence and shorter cancer specific survival in luminal B LN+ breast cancer patients. *Oncotarget* 2018;9:15228.
44. Daves MH, Hilsenbeck SG, Lau CC, et al. Meta-analysis of multiple microarray datasets reveals a common gene signature of metastasis in solid tumors. *BMC medical genomics* 2011;4:56.
45. Ezawa I, Sawai Y, Kawase T, et al. Novel p53 target gene FUCA 1 encodes a fucosidase and regulates growth and survival of cancer cells. *Cancer science* 2016;107:734-745.
46. Cheng T-C, Tu S-H, Chen L-C, et al. Down-regulation of  $\alpha$ -L-fucosidase 1 expression confers inferior survival for triple-negative breast cancer patients by modulating the glycosylation status of the tumor cell surface. *Oncotarget* 2015;6:21283.
47. Zou JX, Duan Z, Wang J, et al. Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. *Molecular Cancer Research* 2014;12:539-549.
48. Park S, Kim H, Ji HW, et al. Cold Atmospheric Plasma Restores Paclitaxel Sensitivity to Paclitaxel-Resistant Breast Cancer Cells by Reversing Expression of Resistance-Related Genes. *Cancers* 2019;11:2011.
49. Yamada KH, Nakajima Y, Geyer M, et al. KIF13B regulates angiogenesis through Golgi to plasma membrane trafficking of VEGFR2. *J Cell Sci* 2014;127:4518-4530.
50. Aurilio G, Disalvatore D, Pruneri G, et al. A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. *European journal of cancer* 2014;50:277-289.
51. Amir E, Miller N, Geddie W, et al. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *Journal of clinical oncology* 2012;30:587.
52. Mao XY, Lee MJ, Zhu J, et al. Genome-wide screen identifies a novel prognostic signature for breast cancer survival. *Oncotarget* 2017;8:14003.
53. Goodwin JM, Svensson RU, Lou HJ, et al. An AMPK-independent signaling pathway downstream of the LKB1 tumor suppressor controls Snail1 and metastatic potential. *Molecular cell* 2014;55:436-450.
54. Chen J, Shen C, Shi J, et al. Knockdown of DIXDC1 inhibits the proliferation and migration of human glioma cells. *Cellular and molecular neurobiology* 2017;37:1009-1019.
55. Zhou S, Shen J, Lin S, et al. Downregulated expression of DIXDC1 in hepatocellular carcinoma and its correlation with prognosis. *Tumour Biol* 2016;37:13607-13616.
56. Wang L, Tan C, Qiao F, et al. Upregulated expression of DIXDC1 in intestinal-type gastric carcinoma: co-localization with beta-catenin and correlation with poor prognosis. *Cancer Cell Int* 2015;15:120.

57. Ma XJ, Dahiya S, Richardson E, et al. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res* 2009;11:R7.
58. Mehrmohamadi M, Liu X, Shestov AA, et al. Characterization of the usage of the serine metabolic network in human cancer. *Cell Rep* 2014;9:1507-1519.
59. Yu M, Ma M, Huang C, et al. Correlation of expression of human arrest-defective-1 (hARD1) protein with breast cancer. *Cancer Invest* 2009;27:978-983.
60. Yu M, Gong J, Ma M, et al. Immunohistochemical analysis of human arrest-defective-1 expressed in cancers in vivo. *Oncol Rep* 2009;21:909-915.
61. Zeng Y, Min L, Han Y, et al. Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis. *Carcinogenesis* 2014;35:2244-2253.
62. El-Arabey AA, Salama SA, Abd-Allah AR. CENP-E as a target for cancer therapy: Where are we now? *Life Sci* 2018;208:192-200.
63. Yardimci H, van Duffelen M, Mao Y, et al. The mitotic kinesin CENP-E is a processive transport motor. *Proc Natl Acad Sci U S A* 2008;105:6016-6021.
64. Kung PP, Martinez R, Zhu Z, et al. Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer. *Mol Cancer Ther* 2014;13:2104-2115.
65. Ohashi A, Ohori M, Iwai K, et al. A Novel Time-Dependent CENP-E Inhibitor with Potent Antitumor Activity. *PLoS One* 2015;10:e0144675.
66. Meister G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* 2013;14:447-459.
67. Fischer KR, Durrans A, Lee S, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* 2015;527:472-476.
68. Voller D, Reinders J, Meister G, et al. Strong reduction of AGO2 expression in melanoma and cellular consequences. *Br J Cancer* 2013;109:3116-3124.
69. Zhang H, Wang Y, Dou J, et al. Acetylation of AGO2 promotes cancer progression by increasing oncogenic miR-19b biogenesis. *Oncogene* 2019;38:1410-1431.
70. Casey MC, Prakash A, Holian E, et al. Quantifying Argonaute 2 (Ago2) expression to stratify breast cancer. *BMC Cancer* 2019;19:712.
71. Conger AK, Martin EC, Yan TJ, et al. Argonaute 2 Expression Correlates with a Luminal B Breast Cancer Subtype and Induces Estrogen Receptor Alpha Isoform Variation. *Noncoding RNA* 2016;2.
72. Adams BD, Claffey KP, White BA. Argonaute-2 expression is regulated by epidermal growth factor receptor and mitogen-activated protein kinase signaling and correlates with a transformed phenotype in breast cancer cells. *Endocrinology* 2009;150:14-23.
73. Logan N, Graham A, Zhao X, et al. E2F-8: an E2F family member with a similar organization of DNA-binding domains to E2F-7. *Oncogene* 2005;24:5000-5004.
74. Lv Y, Xiao J, Liu J, et al. E2F8 is a Potential Therapeutic Target for Hepatocellular Carcinoma. *J Cancer* 2017;8:1205-1213.
75. Kim LK, Park SA, Eoh KJ, et al. E2F8 regulates the proliferation and invasion through epithelial-mesenchymal transition in cervical cancer. *Int J Biol Sci* 2020;16:320-329.
76. Ye L, Guo L, He Z, et al. Upregulation of E2F8 promotes cell proliferation and tumorigenicity in breast cancer by modulating G1/S phase transition. *Oncotarget* 2016;7:23757-23771.
77. Li Y, Huang J, Yang D, et al. Expression patterns of E2F transcription factors and their potential prognostic roles in breast cancer. *Oncol Lett* 2018;15:9216-9230.
78. Iino K, Mitobe Y, Ikeda K, et al. RNA-binding protein NONO promotes breast cancer proliferation by post-transcriptional regulation of SKP2 and E2F8. *Cancer Sci* 2020;111:148-159.
79. Cai Y, Mei J, Xiao Z, et al. Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. *Hereditas* 2019;156:20.

80. Mansouri N, Movafagh A, Sayad A, et al. Targeting of BUB1b Gene Expression in Sentinel Lymph Node Biopsies of Invasive Breast Cancer in Iranian Female Patients. *Asian Pac J Cancer Prev* 2016;17:317-321.
81. Zhao Y, Hu J, Miao G, et al. Transmembrane protein 208: a novel ER-localized protein that regulates autophagy and ER stress. *PLoS One* 2013;8:e64228.
82. Lei P, Xiao Y, Li P, et al. hSnd2/TMEM208 is an HIF-1alpha-targeted gene and contains a WH2 motif. *Acta Biochim Biophys Sin (Shanghai)* 2020.
83. Ren J, Guo H, Wu H, et al. GPER in CAFs regulates hypoxia-driven breast cancer invasion in a CTGF-dependent manner. *Oncol Rep* 2015;33:1929-1937.
84. Deng S, Zhou H, Xiong R, et al. Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast Cancer Res Treat* 2007;104:21-30.
85. Li Y, Huang J, Zeng B, et al. PSMD2 regulates breast cancer cell proliferation and cell cycle progression by modulating p21 and p27 proteasomal degradation. *Cancer Lett* 2018;430:109-122.
86. Tomida S, Yanagisawa K, Koshikawa K, et al. Identification of a metastasis signature and the DLX4 homeobox protein as a regulator of metastasis by combined transcriptome approach. *Oncogene* 2007;26:4600-4608.
87. Lu C, Zhang J, He S, et al. Increased alpha-tubulin1b expression indicates poor prognosis and resistance to chemotherapy in hepatocellular carcinoma. *Dig Dis Sci* 2013;58:2713-2720.
88. Shekhar MP, Lyakhovich A, Visscher DW, et al. Rad6 overexpression induces multinucleation, centrosome amplification, abnormal mitosis, aneuploidy, and transformation. *Cancer Res* 2002;62:2115-2124.
89. Haynes B, Zhang Y, Liu F, et al. Gold nanoparticle conjugated Rad6 inhibitor induces cell death in triple negative breast cancer cells by inducing mitochondrial dysfunction and PARP-1 hyperactivation: Synthesis and characterization. *Nanomedicine* 2016;12:745-757.
90. Somasagara RR, Spencer SM, Tripathi K, et al. RAD6 promotes DNA repair and stem cell signaling in ovarian cancer and is a promising therapeutic target to prevent and treat acquired chemoresistance. *Oncogene* 2017;36:6680-6690.
91. Cabioglu N, Yazici MS, Arun B, et al. CCR7 and CXCR4 as novel biomarkers predicting axillary lymph node metastasis in T1 breast cancer. *Clinical Cancer Research* 2005;11:5686-5693.
92. Shaashua L, Shabat-Simon M, Haldar R, et al. Perioperative COX-2 and  $\beta$ -adrenergic blockade improves metastatic biomarkers in breast cancer patients in a phase-II randomized trial. *Clinical Cancer Research* 2017;23:4651-4661.
93. Millar E, Graham PH, O'Toole SA, et al. Prediction of local recurrence, distant metastases, and death after breast-conserving therapy in early-stage invasive breast cancer using a five-biomarker panel. *J Clin Oncol* 2009;27:4701-4708.