IDENTIFICATION OF A NOVEL EXPERIMENTAL MODEL TO REVEAL MECHANISMS LEADING TO EPIGENETIC CHANGES AND SUBSEQUENT ACTIVATION OF CANCER TESTIS GENES IN CANCER

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

By

Barış Küçükkaraduman

August 2016

IDENTIFICATION OF A NOVEL EXPERIMENTAL MODEL TO REVEAL MECHANISMS LEADING TO EPIGENETIC CHANGES AND SUBSEQUENT ACTIVATION OF CANCER TESTIS GENES IN CANCER

By Barış Küçükkaraduman

August 2016

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

_____

Ali Osmay Güre (Advisor)

_____

Serkan Göktuna

_____

Sreeparna Banerjee

Approved for Graduate School of Engineering and Science:

_____

Levent Onural
Director of the Graduate School

I dedicated my thesis to my mum, dad and sisters for the endless love they gave, my wife being the meaning of my life.

ABSTRACT

IDENTIFICATION OF A NOVEL EXPERIMENTAL MODEL TO REVEAL MECHANISMS LEADING TO EPIGENETIC CHANGES AND SUBSEQUENT ACTIVATION OF CANCER TESTIS GENES IN CANCER

Barış Küçükkaraduman
Master of Science in Molecular Biology and Genetics
Advisor: Ali Osmay Güre
August 2016

Epigenetic aberrations are frequently observed in cancer. Tumor-suppressor genes are often repressed with anomalous hypermethylation in cancer, while DNA hypomethylation has been identified in repetitive sequences and promoter regions of cancer testis (CT) genes resulting in genomic instability. Although it has been shown that CT genes are often regulated by dissociation of repressive proteins from promoter-proximal regions and epigenetic mechanisms, including DNA methylation, histone methylation and acetylation, the process leading to epigenetic changes and de-repression of CT genes remains largely unknown. This study aimed to reveal molecular mechanisms which may have role in coordinating CT gene expression. For this purpose, we designed two groups of experiments. The first was based on extending our previous observations related to two genes (ALAS2, CDR1) which showed inverse expression patterns, compared to CT genes in cancer cell lines. The ex vivo analysis of expression patterns of these genes, however, did not support an inverse relation between their expression and that of CT genes. The second approach was based on categorizing cancer cells into CT-high, CT-intermediate and CT-low groups to define differentially expressed non-CT genes that could help explain mechanisms underlying epigenetic changes and subsequent activation of CT genes. Surprisingly, we could not identify any transcripts that differentially expressed between these subgroups. We therefore, hypothesized that non-overlapping and distinct mechanisms could be involved in the upregulation of CT genes in different tumors. As our earlier work suggested a relationship between epithelial to mesenchymal transition (EMT) and CT expression we asked if an EMT based classification could help elucidate these mechanisms. Indeed, differential genes and differentially activated signaling pathways were discovered when cancer cells were first grouped by their EMT status. This helped us identify candidate

proteins (BMI1, PCGF2, RB1 and RBL1) and pathways including MAPK/ERK and PTEN/PI3K pathways which can coordinate CT gene expression in cancer. Thirdly, we investigated clinical relevance of high CT gene expression in triple negative breast cancer by attempting to correlate this with drug sensitivity. Drug sensitivity against panobinostat showed correlation with CT gene expression. In summary, this study suggests new approaches to elucidate mechanisms which coordinate epigenetic aberrations in cancer and how these can be utilized for cancer therapy.

Keywords: Cancer testis genes, DNA methylation, epithelial and mesenchymal phenotype, panobinostat

ÖZET

KANSERDE EPİGENETİK DEĞİŞİMLERE VE SONRASINDA KANSER TESTİS GENLERİNİN AKTİVASYONUNA SEBEP OLAN MEKANİZMALARIN ORTAYA ÇIKARILMASI İÇİN YENİ DENEYSEL MODELLERİN TANIMLANMASI

Barış Küçükkaraduman
Moleküler Biyoloji ve Genetik, Yüksek Lisans
Tez Danışmanı: Ali Osmay Güre
Ağustos 2016

Normal olmayan epigenetik değişimler kanserde sıkça görülür. Kanserde, tümör baskılayıcı genler çoğu kez normal olmayan hipermetilasyon ile susturulurken, tekrar eden sekansların ve kanser testis (KT) genlerinin promoter bölgelerinin hipometilasyon uğradığı belirlenmiştir. KT genlerinin, promotere yakın bölgelerdeki baskılayıcı proteinlerin ayrılması ve DNA metilasyonu, histon metilasyon ve asetilasyonu gibi epigenetik mekanizmalarla kontrol edildiği gösterilmesine rağmen bu epigenetik değişimlere ve KT genlerinin baskıdan kurtulmasına neden olan süreç büyük oranda netlik kazanmamıştır. Bu çalışma KT gen ifadesini koordine eden moleküler mekanimların ortaya çıkarılmasını amaçlamaktadır. Bu amaçla, iki grup deney dizayn edilmiştir. İlki, KT genleri ile karşılaştırıldığında ters ifade düzenine sahip olan iki gen (ALAS2, CDR1) ile ilgili önceki gözlemlerimizin geliştirilmesine dayanır. Bu genlerin gen ifade düzeninin ex vivo analizi, bu genlerin ifadesinin KT genleri ile ters ilişkili olduğunu desteklememiştir. İkinci yaklaşım, KT geni olmayan fakat epigenetik değişimleri ve sonrasındaki KT genlerinin aktivasyonunu sağlayan mekanizmaları açıklamamıza yardımcı olabilecek, gen ifadesinde farklılık gösteren genlerin bulunması için kanser hücrelerinin KT-Yüksek, KT-Ortave KT-Düşük olarak kategorize edilmesine dayanmaktadır. Beklenmedik şekilde, bu gruplar arasında gen ifadesinde farklılık gösteren ortak genler bulunumamıştır. Bu yüzden, birbiri ile örtüşmeyen farklı mekanizmaların, farklı tümörlerdeki KT genlerinin aktivasyonunda ilişkili olabileceği varsayımında bulunduk. Bir önceki çalışmamızda epitel-mezenkimal geçiş (EMG) ile KT gen ifadesi ilişkisini göz önünde bulundurarak, EMG'ye dayalı bir sınıflandırmanın KT gen ifadesini kontrol eden mekanizmayı açıklayıp açıklayamayacağını araştırdık. Kanser hücreleri EMG statülerine göre gruplandırıldıklarında, bazı genlerin ifadelerinin ve sinyal iletim yolaklarının farklılaşarak aktive olduklarını keşfettik. Bu yaklaşım, bizim KT gen ifadesinin kontrolünde rol oynayabilecek bazı aday protein (BMI1, PCGF2, RB1 ve RBL1) ve MAPK/ERK ve PTEN/PI3K gibi sinyal iletim yolaklarını bulmamızı sağladı. Üçüncü olarak, üçlü negatif meme kanserindeki yüksek KT gen ifadesinin klinik bir önemi olup olmadığını ilaç hassiyetleri ile korelasyonuna bakarak araştırdık. Panobinostat hassasiyeti, KT gen ifadesi ile pozitif korelasyon gösterdi. Özetle, bu çalışma kanserdeki epigenetik anormallikleri koordine eden mekanizmaları ve bunların kanser tedavisi için nasıl kullanılabileğini ortaya koyan yeni yaklaşımlar içermektedir.

Anahtar kelimeler: Kanser testis genleri, DNA metilasyonu, epitel-mezenkimal geçiş, panobinostat

Acknowledgements

I would like to express my sincere gratitude and appreciation to my advisor Dr. Ali Osmay Güre for his guidance in this work. His wide knowledge, support and inspiring suggestions helped me improve myself in the field of science.

For being my thesis committee members and their important critics, I would like to thank Assoc. Prof. Dr. Sreeparna Banerjee and Asst. Prof. Dr. Serkan Göktuna.

I would like to acknowledge Murat İşbilen, Seçil Demirkol for their remarkable efforts on bioinformatics in this projects.

My greatest gratitude goes to my friend Alper Poyraz for his endless support. I am also deeply thankful to current and former members of AOG group; Şükrü Atakan, Sinem Yılmaz-Özcan, Seçil Demirkol, Waqas Akbar, Kerem Mert Şenses, Mehdi Ghasemi since we altogether generated a positive and social work environment.

I am especially grateful to my friends, Erol Eyüpoğlu, Ali Can Savaş, Özge Saatçi, Merve Mutlu, Umar Raza, Pelin Ersan and Hilal Bal for their companionship.

I would like to express my very great appreciation to all Bilkent MBG family. Füsun Hanım, Yavuz Abi, Abdullah Amca, Ümmühan Hanım made this department as a home for all of us.

Above all, I am very grateful to my mother, father and sisters for being beside me and believing me in all stages of my educational life. I have been the most lucky person to have such strong, supporting, loving and patient wife, Büşrü Küçükkaraduman.

I was supported by TÜBİTAK 2211 scholarship program during my study, thereby I would like to thank TÜBİTAK for giving me this opportunity.

Table of Contents

List of Figures

List of Tables

Abbreviations

5-AZA	5-aza-2'-deoxycytidine
ALAS2	5'-Aminolevulinate Synthase 2
ATP	Adenosine triphosphate
BMI1	Polycomb Group RING Finger Protein 4
BORIS	CCCTC-binding factor like
CCLE	Cancer Cell Line Encyclopedia
CDH1	Cadherin-1
CDR1	Cerebellar degeneration related protein 1
CDX2	Caudal Type Homeobox 2
CGAP	Cancer Genome Anatomy Project
CGP	Cancer Genome Project
cDNA	Complementary deoxyribonucleic acid
CSC	Cancer stem cell
CT	Cancer testis
CTCF	CCCTC-binding factor
ddH_2O	Double distilled water
DET	Differentially expressed transcripts
DMEM	Dulbecco's Modified Eagle Medium
EMEM	Eagle's Minimum Essential Medium
EMT	Epithelial-mesenchymal transition
ERK	Mitogen-activated protein kinase
FBS	Fetal bovine serum
FN	Fibronectin
GAGE	G Antigen
GSEA	Gene set enrichment analysis
HDACi	Histone deacetylase inhibitor
HP1	Heterochromatin protein 1
KMT6	Enhancer Of Zeste Homolog 2
KRAS	Kirsten rat sarcoma viral oncogene homolog

IC50	Inhibitory concentration 50%
LINE1	Long interspersed nuclear elements
MAGE	Melanoma Antigen Family
MAP2K1	Mitogen-Activated Protein Kinase Kinase 1
MET	Mesenchymal-epithelial transition
mRNA	Messenger ribonucleic acid
NY-ESO-1	New York Esophageal Squamous Cell Carcinoma 1
OD	Optical density
PAGE	Prostate associated gene family
PBS	Phosphate buffered saline
PCGF2	Polycomb Group Ring Finger 2
PI3K	Phosphoinositide 3-kinase
PRC2	Polycomb Repressive Complex 2
PTEN	Phosphatase and tensin homolog
RB1	Retinoblastoma 1
RBL1	Retinoblastoma-Like 1
RPMI	Roswell Park Memorial Institute medium
RNA	Ribonucleic acid
SAHA	Suberoyl hydroxamic acid, Vorinostat
SPANX	Sperm protein associated with the nucleus, X chromosome family
SSX	Sarcoma, synovial, X-chromosome-related gene family
TET2	Tet Methylcytosine Dioxygenase 2
TNBC	Triple negative breast cancer
TSA	Trichostatin A
VIM	Vimentin

1. INTRODUCTION

1.1 Cancer Testis Genes

Cancer-Testis (CT) Genes are group of genes which show selective expression patterns [1]. Their expression is restricted to adult testis germ cells among healthy tissues and they are reactivated in various tumor types [2]. Additionally, CT gene expression is also observed in fetal ovary and placenta [3]. Epigenetic changes which result in anomalous gene expression patterns are recognized as a hallmark of cancer. Tumorigenesis is promoted and driven by group of genes; some of them are called epigenetic-drivers. Aberrant expression patterns in cancer are thought to be crucial characteristics of epigenetic-drivers of tumorigenesis [4-8]. The re-activation of CT genes in cancer is thought to involve similar epigenetic changes that are observed in gametogenesis and tumorigenesis [9]. Therefore, studying epigenetic mechanisms which control CT gene expression may provide a better understanding of tumorigenesis. Despite the fact that CT gene expression occurs irrespective of tissue of origin in cancer, it is not observed in all samples of a given cancer type. This selective expression pattern of CT genes is a valuable model to examine epigenetic alterations behind aberrant gene expression and complex regulations in carcinogenesis.

To date, more than 200 CT genes have been reported [10]. As most CT gene products are capable of inducing autologous anti-tumor immune responses, autologous typing was the first methodology to identify CT antigens [11]. With the development of microarray and next-generation sequencing technologies, new CT antigens were identified recently by using multiple databases [12]. Studying CT gene expression *in silico* is challenging as most CT genes exist as families with highly homologous members which are difficult to distinguish in microarray or RNA sequencing based methods. Additionally, there is lack of *in vitro* experimental models in which epigenetic regulations of CT genes can be studied.

CT genes are expressed in various cancer types which can thus be designated as "CT-rich" or "CT-poor". CT-rich tumors include melanomas, hepatocellular carcinomas, chondrosarcomas, non-small lung, ovarian and bladder cancers. Tumors originated from breast and prostate have moderate expression of CT genes. Tumors defined as "CT-poor" include renal, pancreatic and colon cancers, and hematological malignancies [13, 14]. There are different classifications of CT genes based on their chromosomal localizations, "CT-X" genes and "non-X CT" genes [1], or based on their mRNA expression spectrum in normal tissues, "testis/brain-selective", "testis-restricted" and "testis-selective" [15].

Although CT genes have heterogeneous expression among different cancer types, it has been shown that CT genes are coordinately expressed in non-small lung cancer as well as in other cancers [16]. This is an important observation suggesting that mechanism of CT gene expression regulation is controlled by common mechanisms [17].

The common mechanism which controls CT gene expression is thought to be mainly epigenetic in nature [13]. Studies show that healthy tissues lacking CT gene expression have DNA hypermethylation in their promoter region. In contrast, CT expressing cancers and testis tissue show DNA hypomethylation in promoter regions of CT genes [18]. This fact is also observed in 5-aza-2'-deoxycytidine (5-AZA, a DNA hypomethylating agent) induced cell lines in which 5-AZA treated cell lines have CT gene upregulation with DNA hypomethylation in their promoter region, suggesting that DNA methylation is a crucial epigenetic change in controlling expression of CT genes [19]. DNA hypomethylation in promoter region of CT genes is observed in parallel to LINE1 repeat demethylation indicating that global DNA hypomethylation is a leading epigenetic mechanism in re-activation of CT genes in cancer[20-23]. Intratumor and intertumor heterogeneity of CT gene expression is associated with promoter region-specific and global DNA methylation status in tumors as well [19, 24], suggesting the mechanisms that control methylation status of CT genes are likely to be complex. In

other studies, it has been shown that histone acetylation results in upregulation of CT gene expression [25, 26]. In another study, findings suggest that CT gene expression is both associated with DNA methylation and EZH2-H3K27Me3 status. It was shown that siRNA silencing of EZH2, in combination with 5-AZA and TSA treatment, induced strong activation of GAGE family CT genes while 5-AZA and TSA treatment alone resulted in weak induction [27]. This is concordant with our results where we showed EZH2 dissociation from CT gene promoters during Caco-2 differentiation resulting in upregulation of PAGE2 and SPANXB [36]. KMT6 mediates trimethylation of H3K27 which is associated with polycomb repressive complex 2 (PRC2) binding. KDM1 and KDM5B mediate demethylation of H3K4 which is associated repressed genes. Inhibition of these proteins was shown to improve the effect of DNA hypomethylating agent 5-AZA in terms of CT gene upregulation [28-30]. In summary, repressor complexes, histone acetylation, histone and DNA methylation, and associated proteins were shown to play a role in the regulation of CT gene expression [28-30][28-30][28-30]. Transcription factors have also shown to be important regulators of CT gene expression. In lung cancer cells, promoter region of NY-ESO-1 is occupied with transcription factors CTCF and BORIS resulting in de-repression of this gene [31]. Re-activation and de-repression are also observed in conditional expression of BORIS in healthy cells resulting in DNA hypomethylation [32]. Our studies also suggested that CT genes in a defined region are epigenetically activated, while other regions of genome have different epigenetic changes which repress genes in the region. So, genomic locations on which CT genes are found have clear boundaries, thus these regions and outside of the regions are controlled by different epigenetic mechanisms.

Previously, members of our group hypothesized that if CT genes were co-regulated, that there could be genes whose expression followed the exact opposite pattern. In this line, we showed that CT genes, PAGE2B and SPANXB, are up-regulated while two genes proximal to these (identified by an *in silico* method involving the CGAP database), ALAS2 and CDR1, are down-regulated in cancer while the opposite expression pattern is observed for normal tissues (Figure 1.1, 1.2, 1.3, 1.4).

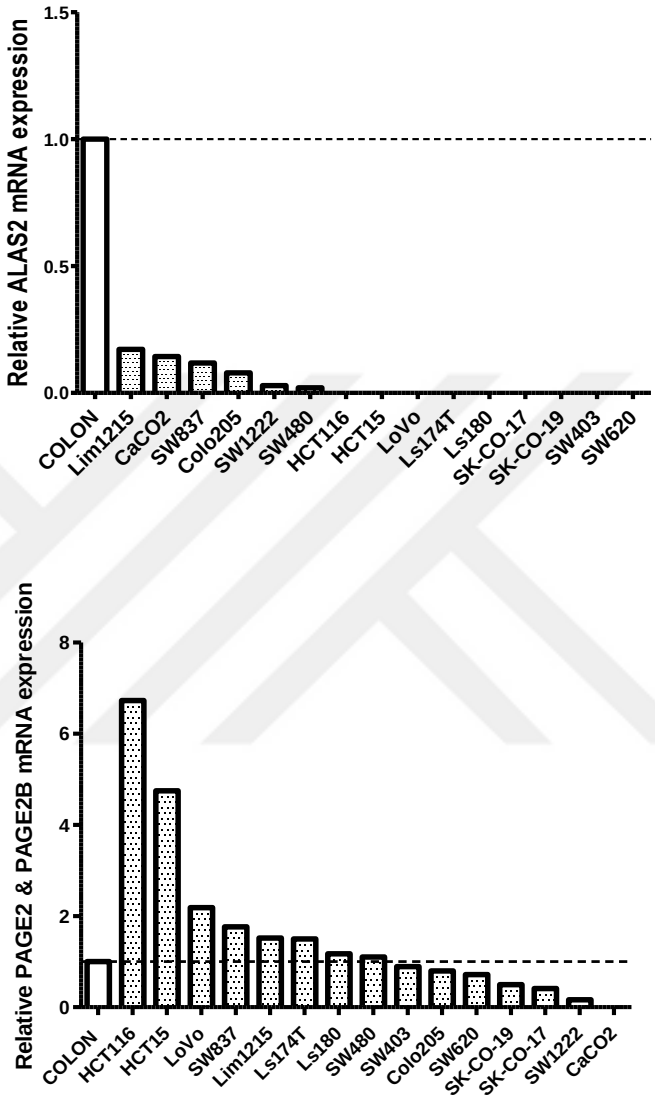


Figure 1.1: ALAS2 and PAGE2&2B mRNA expression in healthy colon and colon cancer cell lines. The significant down-regulation in ALAS2 (50 kb away from PAGE2 gene) expression was observed in a panel of colon cancer cell lines compared to normal tissue while PAGE2&2B mRNA expression was significantly up-regulated in cancer cell lines compared to normal. GAPDH gene expression was used as endogenous control. These data adapted from our previous studies.

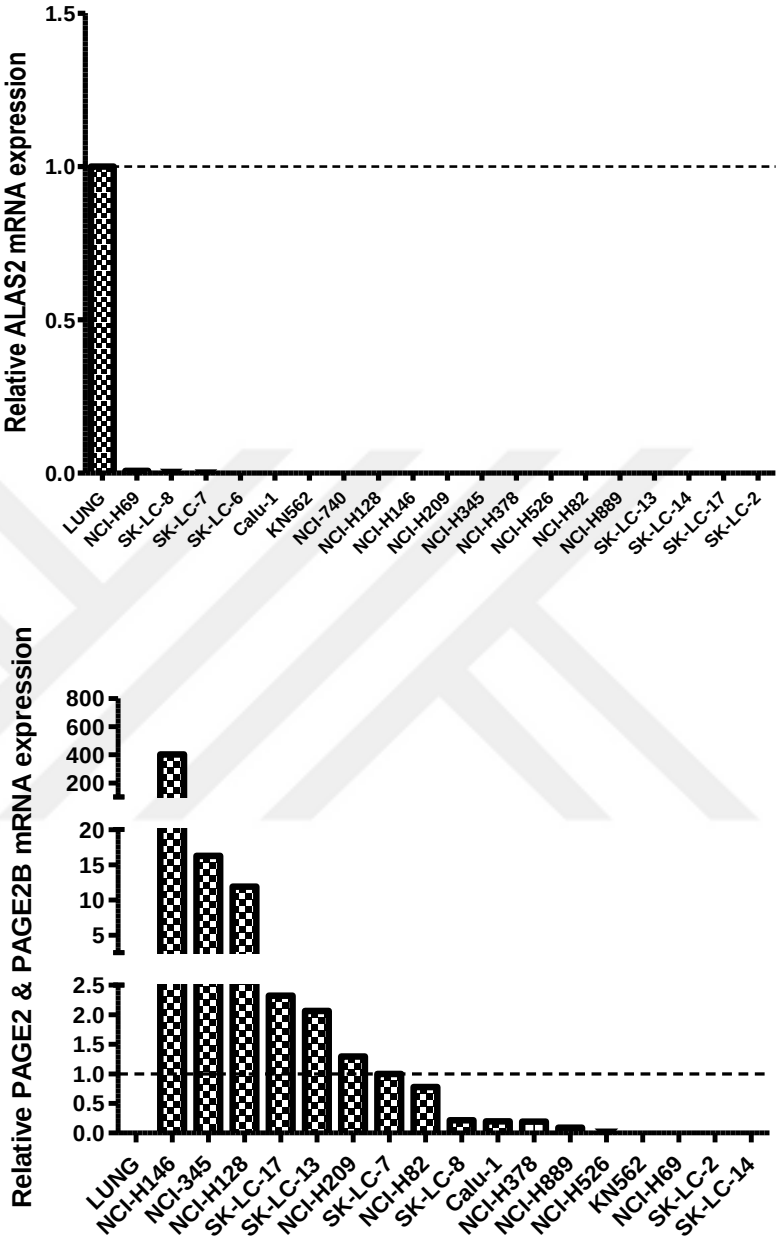


Figure 1.2: ALAS2 and PAGE2&2B mRNA expression in healthy lung and lung cancer cell lines. The significant down-regulation in ALAS2 (50 kb away from PAGE2 gene) expression was observed in a panel of lung cancer cell lines compared to normal tissue while PAGE2&2B mRNA expression was significantly up-regulated in cancer cell lines compared to normal. GAPDH gene expression was used as endogenous control. These data adapted from our previous studies.

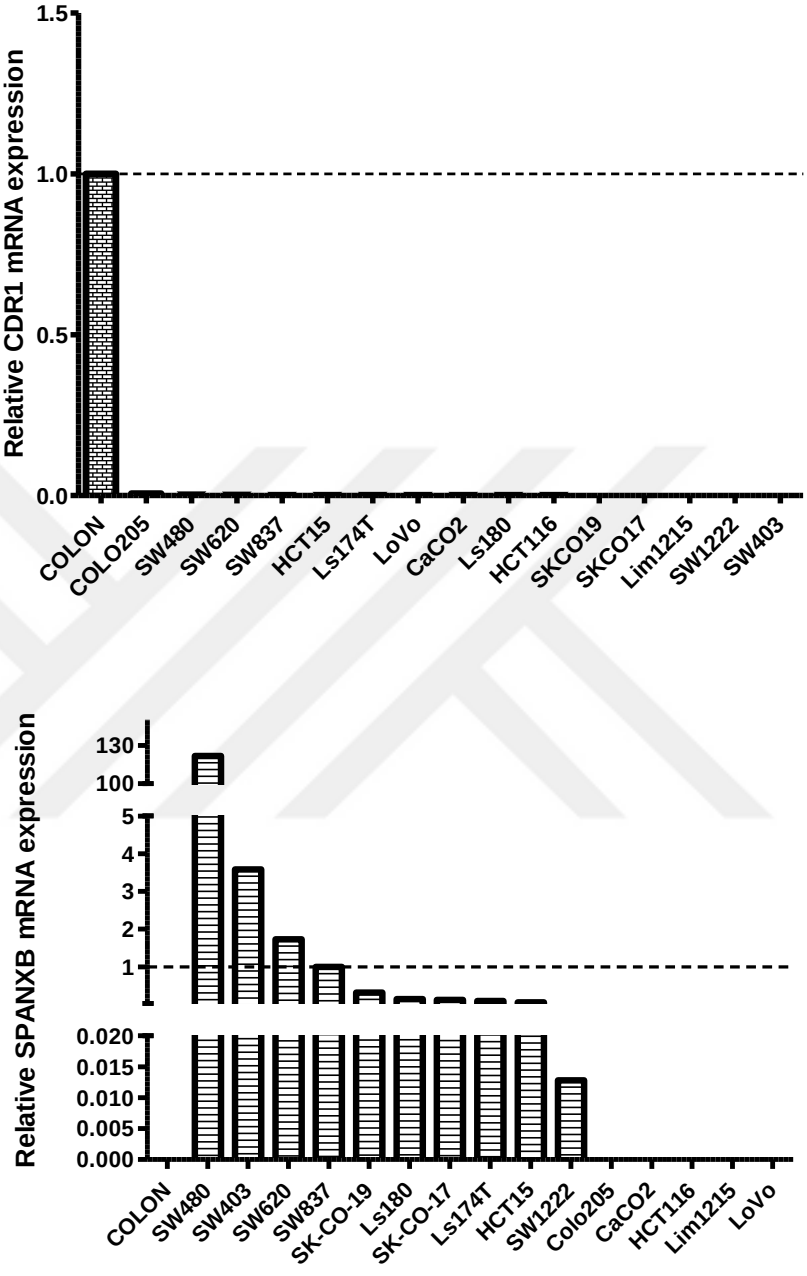


Figure 1.3: CDR1 and SPANXB mRNA expression in healthy colon and colon cancer cell lines. The significant down-regulation in CDR1 (50 kb away from SPANXB gene) expression was observed in a panel of colon cancer cell lines compared to normal tissue while SPANXB mRNA expression was significantly up-regulated in cancer cell lines compared to normal. GAPDH gene expression was used as endogenous control. These data adapted from our previous studies.

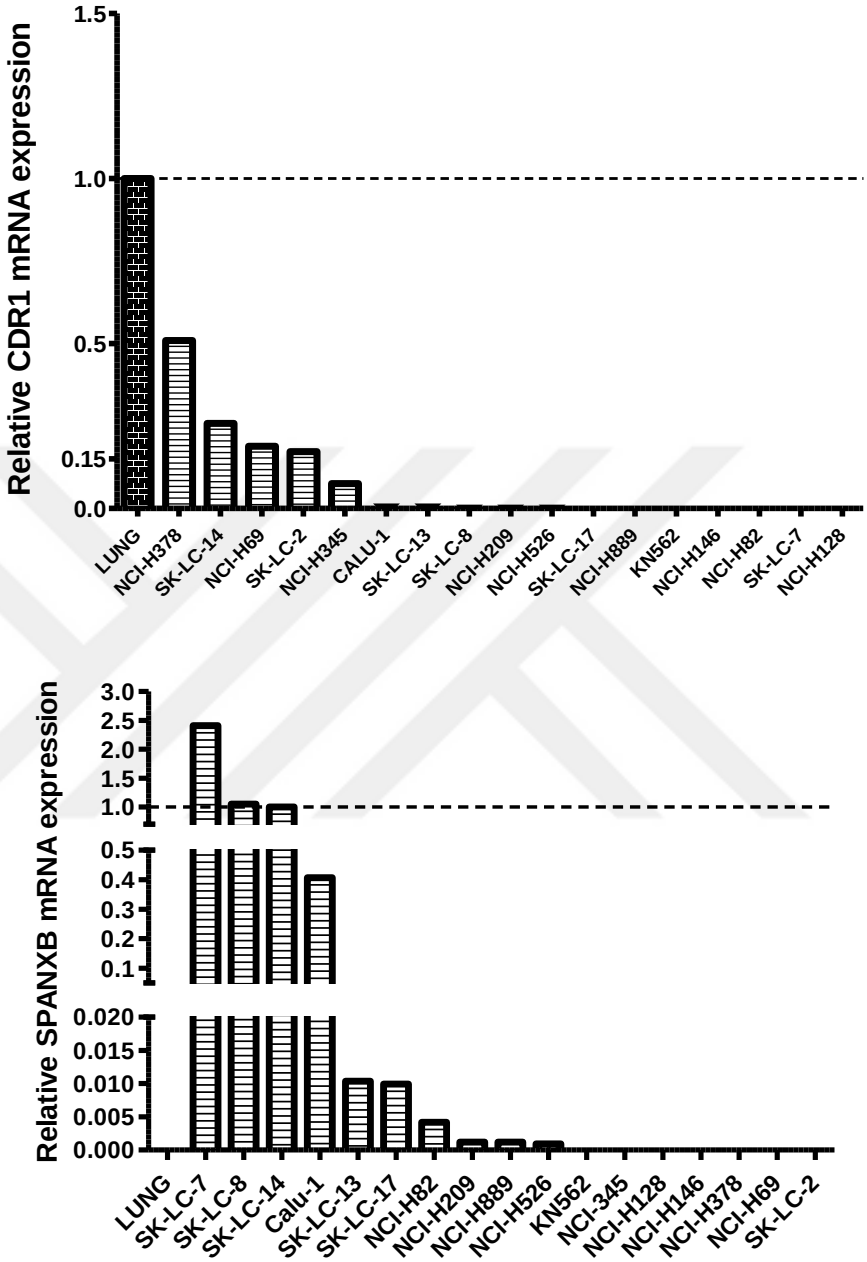


Figure 1.4:CDR1 and SPANXB mRNA expression in healthy lung and lung cancer cell lines. The significant down-regulation in CDR1 (50 kb away from SPANXB gene) expression was observed in a panel of colon cancer cell lines compared to normal tissue while SPANXB mRNA expression was significantly up-regulated in cancer cell lines compared to normal. GAPDH gene expression was used as endogenous control. These data adapted from our previous studies.

1.2 Cancer Testis Antigens and Epithelial-Mesenchymal Transition

CT gene expression status during epithelial-to-mesenchymal transition (EMT) is very controversial. Some reports indicate that CT genes are highly expressed in cells with an epithelial phenotype (or differentiated) while others report that there is CT gene upregulated in cells with mesenchymal phenotype (or a stem-cell like). For instance, SSX, a CT gene family, is highly expressed in human mesenchymal stem cells and was shown to co-localize with mesenchymal marker gene, Vimentin [33]. Similarly, in another study, MCF-7, a breast cancer cell line, which overexpresses SSX showed repressed E-Cadherin, an epithelial marker gene [34]. CT45 which is a testis-restricted CT gene has been identified with aberrant expression in epithelial cancers [35]. In our recent findings, in contrast to findings summarized above, we showed that mesenchymal-to epithelial transition (MET) resulted in upregulation of some CT genes. In this study, we used Caco-2 colon cancer cell line to examine CT gene expression during mesenchymal-to epithelial transition. When Caco-2 cells reach confluency, they start to differentiate with phenotypic changes. In this model, CT gene expression was measured at post-confluence, day 10, 20 and 30. SPANXB and PAGE2 genes were upregulated during differentiation (Figure 1.5). If cells at day 20 were detached and seeded resulting in de-differentiation, downregulation of SPANXB and PAGE2 genes were observed together with upregulation of mesenchymal markers. In the same study, we showed that there is an increase in TET2 expression and 5-hydroxymethylation levels in promoter regions of SPANXB and PAGE2 genes concordant with MET (Figure 1.6). Dissociation of both PRC2 and HP1 repressor proteins (Figure 1.7) from promoter regions of these genes was also observed during MET [36]. Although this study helped identify novel epigenetic changes which relate to the activation of these genes, we still don't have a complete understanding of how these genes are selectively activated and which mechanisms lead to these epigenetic changes that result in CT gene expression.

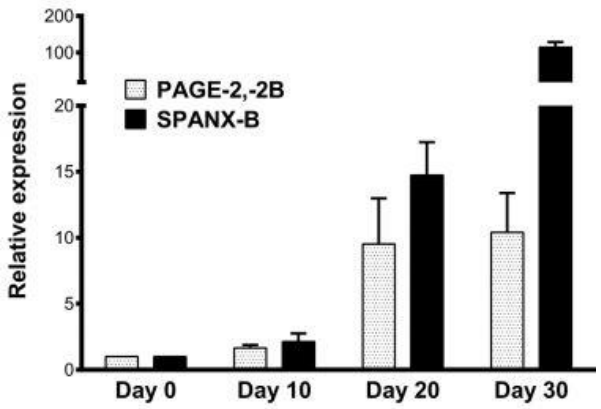


Figure 1.5:Expression levels of PAGE-2,-2B and SPANX-B during Caco-2 spontaneous differentiation, at days 0, 10, 20 and 30. Significant upregulation was detected in PAGE-2,-2B and SPANX-B gene expression during mesenchymal-to-epithelial transition. The figure retrieved from our previous study [36].

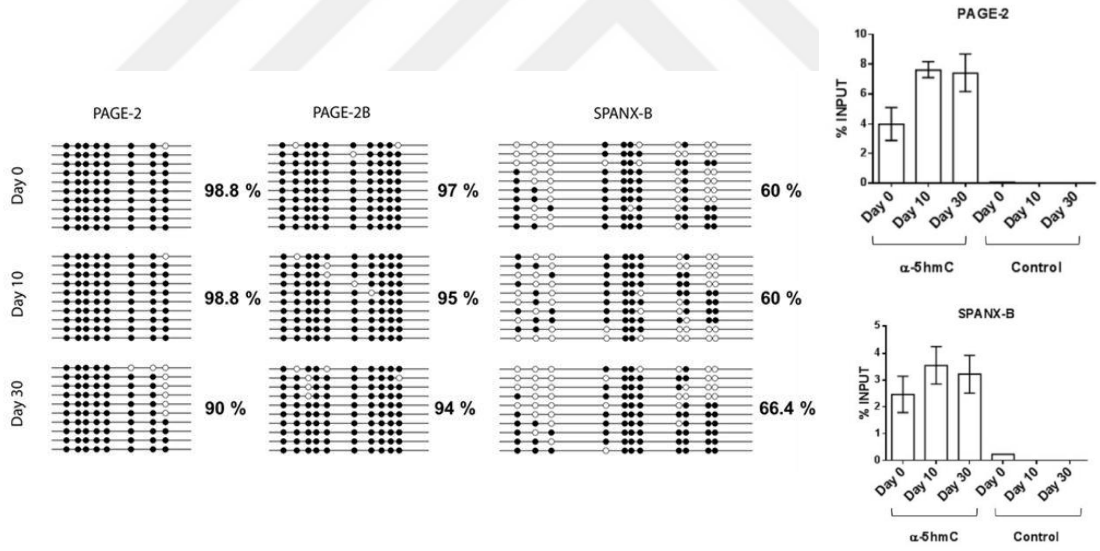


Figure 1. 6: DNA methylation and hydroxymethylation levels during spontaneous differentiation of Caco-2 cells. Hydroxymethylation levels (at right) in promoter region of CT genes, PAGE-2 and SPANX-B, increases during mesenchymal to epithelial transition while promoter proximal DNA regions of CT genes are heavily hypermethylated (at left). The figure retrieved from our previous study [36].

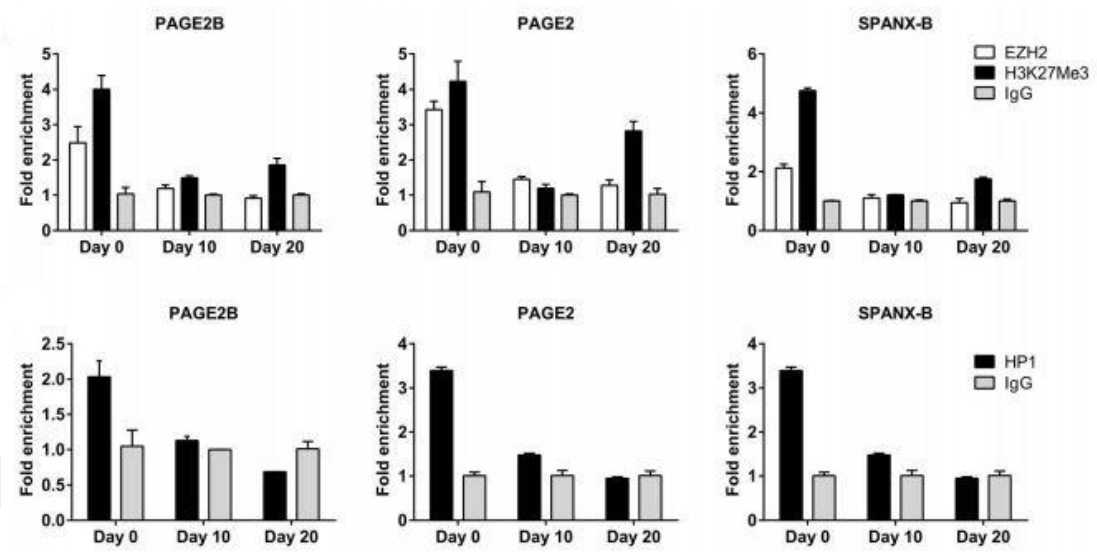


Figure 1.7: Chromatin modifications in PAGE2, 2B and SPANXB during Caco-2 differentiation. Heterochromatin protein 1 (HP1) and polycomb repressive complex 2 protein EZH2 are dissociated from promoter-proximal regions of CT genes, PAGE-2 and SPANX-B, and H3K27Me3 levels were decreased during mesenchymal to epithelial transition. The figure retrieved from our previous study [36].

1.3 Potential Clinical Value of CT Gene Expression

Various tumor associated antigens were identified in several studies and this leads to development of new approaches for targeted immunotherapy of cancers [37]. One of these targets for immunotherapy is cancer testis antigens because of their particular expression pattern in healthy and cancerous tissues [1]. Most of studies about clinical significance of cancer testis gene expression suggest correlation of CT expression with worse prognosis while CT expression has been less frequently linked to improved outcome in different cancer types [38-42]. However, clinical significance of CT gene expression as being chemosensitivity biomarker in cancers remains largely unknown. In this study, we aimed to show potential clinical value of CT gene expression in breast cancer by attempting to correlate CT gene expression with drug response data from CCLE and CGP.

1.3.1 Breast Cancer

Breast cancer is a complex disease which is the most frequently diagnosed cancer and the most common cause of death from cancer in women after lung cancer. Risk factors are divided into three groups, potentially modifiable factors, non-modifiable factors and reproductive factors. Long-term heavy smoking, alcohol consumption, weight gain after 18 in years of age, obesity, being on menopausal hormone therapy, and physical inactivity are called modifiable factors which are associated with increased breast cancer risk [43-47]. High dose radiation to the chest, inherited mutations in risk factor genes, older age, history of hyperplasia, ductal or lobular carcinoma, and type 2 diabetes are classified as non-modifiable factors [48-52]. Reproductive factors include high levels of sex hormones, aberrant menstrual history, never having children, and having one's first child after 30 years of age [53-57]. Treatment strategies for breast cancer depend on tumor characteristics. Breast conserving surgery or removal of all breast tissue is the usually commonly performed. Treatment options also involve targeted therapy, hormonal therapy, radiation therapy and chemotherapy. Despite there being advanced strategies to treat breast cancer, resistance to treatment is very common [58-62].

1.3.2 Classification of Breast Cancer and Breast Cancer Cell Lines

Breast tumors and frequently used breast cancer cell lines are mostly classified in various studies based on their histopathology, grade, stage, hormone receptor status, gene expression and mutation status [63-67]. In classification of breast cancer cell lines, genome copy number and transcriptional profiles of cell lines and primary breast tumors were compared [68, 69]. Breast cancer cell lines were classified into Luminal, Basal A and Basal B while intrinsic subtypes of breast cancer has been identified as Luminal A, Luminal B, HER2-enriched, Basal-like and Normal-like [66, 68, 70]. These classifications are mostly based on differences in gene expression analyses which also provide valuable insights into breast cancer complexity. Classification of breast cancer cell lines revealed that cell lines mirror most of primary breast tumors with some

differences. For instance, Basal-like intrinsic subtype of breast tumors has been subdivided into two termed Basal A and Basal B. It has been shown that Vimentin-positive Basal B subtype is clearly distinct from Basal A subtype exhibiting a stem cell-like expression pattern. On the other hand, Basal A subtype reflects the features of the clinical triple-negative tumor type. Luminal cluster of cancer cell lines mirror Luminal A and Luminal B intrinsic subtypes of breast primary tumors [66, 70-72].

Table 1.1: Intrinsic subtypes of breast cancer cell lines

Cell line	Gene cluster	Cell line	Gene cluster
600MPE	Lu	MCF7	Lu
AU565	Lu	MDAMB134VI	Lu
BT20	BaA	MDAMB157	BaB
BT474	Lu	MDAMB175VII	Lu
BT483	Lu	MDAMB231	BaB
BT549	BaB	MDAMB361	Lu
CAMA1	Lu	MDAMB415	Lu
HBL100	BaB	MDAMB435	BaB
HCC1007	Lu	MDAMB436	BaB
HCC1143	BaA	MDAMB453	Lu
HCC1187	BaA	MDAMB468	BaA
HCC1428	Lu	SKBR3	Lu
HCC1500	BaB	SUM1315MO2	BaB
HCC1569	BaA	SUM149PT	BaB
HCC1937	BaA	SUM159PT	BaB
HCC1954	BaA	SUM185PE	Lu
HCC202	Lu	SUM190PT	BaA
HCC2157	BaA	SUM225CWN	BaA
HCC2185	Lu	SUM44PE	Lu
HCC3153	BaA	SUM52PE	Lu
HCC38	BaB	T47D	Lu
HCC70	BaA	UACC812	Lu
HS578T	BaB	ZR751	Lu
LY2	Lu	ZR7530	Lu
MCF10A	BaB	ZR75B	Lu
MCF12A	BaB		

Lu: Luminal
BaA: Basal A
BaB: Basal B

1.3.3 CT Gene Expression in Breast Cancer

CT genes are moderately expressed in breast cancer [2]. Additionally, there are a number studies on the clinical relevance of CT gene expression in breast cancer. A recent finding suggests that AKAP3 expression in breast tumors and normal adjacent tissues may be good predictor of prognosis [73]. Another study examined the expression of NY-ESO-1 in 623 breast tumors, suggesting that NY-ESO-1 gene is highly expressed in triple-negative breast cancer and it is related with good prognosis [74]. On the other hand, large scale study by Grigoriadis et al. showed that MAGEA family of CT genes and NY-ESO-1 are highly expressed in ER-negative breast tumors [75]. MAGEA9 expression was also evaluated in invasive ductal breast cancer, and results showed that higher MAGEA9 expression was correlated with poor prognosis [76]. It was also shown that other MAGEA family genes, MAGEA1, MAGEA6 and MAGEA12 are frequently expressed in breast cancer [77] and identified as causal factor in the formation of tamoxifen resistant breast cancer [78]. Therefore CT gene expression has been associated with good and bad prognosis groups in different studies.

1.3.4 Relationship Between CT Gene Expression and Drug Sensitivity

A subtype of breast cancer, triple negative breast cancer (TNBC), is characterized by the lack of estrogen receptor, progesterone receptor expression and lacking amplification of human epithelial growth factor receptor 2. This type of cancer is generally aggressive, and therapeutic approaches are very limited [79, 80]. Studies on CT expression in breast cancer showed that basal-like subtype tumors have a high variation of CT gene expression compared to other subtypes [12, 81]. This heterogeneity could be reflecting distinct cellular pathways being active in CT-High and CT-Low cells, and if this is so then CT gene expression in TNBCs could correlate with drug response.

1.4 Aim and Hypothesis

Since CT gene expression shows variability in cancer, studying CT gene expression can reveal distinct mechanisms active in CT-High and CT-Low cells and the study of these could help understand epigenetic-drivers tumorigenesis. In this line, studying CT gene expression patterns could also help develop new personal therapeutic approaches by predicting drug responses.

In the present study, we aimed to identify an experimental model whose study could reveal molecular mechanisms which may have role in controlling epigenetic changes in promoter regions of CT genes. We assumed that “CT-rich”, “CT-moderate” and “CT-poor” tumors would correspond to cancer cell lines, with CT high, CT intermediate and CT low expression. Analyzing gene expression differences between CT high and low cell line groups, we tried to identify non-CT genes that could be responsible for this difference. Secondly, with the knowledge of the role of EMT in CT gene expression, we performed a similar analysis in cancer cell lines that were first classified according to their epithelial/mesenchymal phenotype. Finally, we studied clinical relevance of CT gene expression as a marker of targeted drug responses in Basal B subtype of breast cancer cell lines.

2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 General Laboratory Reagents and Equipment

Table 2.1: List of laboratory reagents used in this study

Material	Catalog Number	Company
DEPC-Treated Water	AM9920	Ambion by Life Sciences (CA, USA)
TRIzol Reagent	15596026	Ambion by Life Sciences (CA, USA)
RNaseZap RNase Decontamination Solution	AM9782	Ambion by Life Sciences (CA, USA)
DNA AWAY Surface Decontaminant	7010	Molecular BioProducts (CA, USA)
Qubit RNA BR Assay Kit	Q10210	Invitrogen (CA, USA)
Maxima First Strand cDNA Synthesis Kit for RT-qPCR	K1641	Thermo Scientific (IL, USA)
DNA-free DNA Removal Kit	AM1906	Ambion by Life Sciences (CA, USA)
OneTaq Hot Start DNA Polymerase	M0481S	New England BioLabs (MA,USA)
LightCycler 480 SYBR Green I Master	04707516001	Roche Diagnostics (Basel, Switzerland)
Taqman Gene Expression Master Mix	4369016	Applied Biosystems by Life Sciences (CA,USA)
Gene Ruler 100 bp DNA Ladder	SM0241	Thermo Scientific (IL, USA)
Gene Ruler DNA Ladder Mix	SM0333	Thermo Scientific (IL, USA)

Table 2.1: List of instruments used in this study

Instruments	Company
Eppendorf 5810 R Refrigerated Centrifuge	Eppendorf (Hamburg, Germany)
Heat Block	
Qubit Fluorometer	Invitrogen (CA, USA)
NanoDrop One	Thermo Scientific (DE, USA)
Merinton SMA1000 Spectrophotometer	Merinton (MI,USA)
XCell SureLock Mini-Cell Electrophoresis System	Life Sciences (CA, USA)
Mini-Sub Cell GT Horizontal Electrophoresis System	BIO-RAD (CA, USA)
Light Cycler 480 II PCR Machine	Roche (Basel, Switzerland)
Bio-Tek Synergy HT Multi-Mode Microplate Reader	BioTek Instruments (VT, USA)
CO2 Incubator NU-8500 Water Jacket	NuAire (MN,USA)
LabGard NU-425 Class II Biosafety Cabinet	NuAire (MN,USA)

2.1.2 Cell Lines, Culture Materials and Reagents

Breast cancer Basal B subtype cell lines (CAL-51, HBL-100, HCC38, MDA-MB-231, MDA-MB-231, MDA-MB-436 and SUM149PT) and Basal A subtype cell line (BT-20) were obtained from LGC Standards (Middlesex, UK).

Cell culture materials; T-25 and T-75 flask, 100 cm^2 cell culture dishes, 5ml and 10 ml serological pipets were purchased from Costar Corning INC (NY, USA) and 6-well,

12-well, 96-well plates , cryotubes, 10, 20, 200, and 1000 ul filtered tips for micropipettes were purchased from Greiner Bio-One (NC,USA). Cell culture scrapers were purchased from Sarstedt (Numbrecht, Germany).

Table 2.2: List of reagent used in cell culture.

Reagents	Company
RPMI Medium	Capricorn Scientific (Ebsdorfergrund, Germany)
DMEM	Lonza (Basel, Switzerland)
Ham's F12 Medium	Lonza (Basel, Switzerland)
Fetal Bovine Serum	Capricorn Scientific (Ebsdorfergrund, Germany)
Trypsin-EDTA	HyClone (IL, USA)
L-Glutamine	HyClone (IL, USA)
Penicillin-Streptomycin	Capricorn Scientific (Ebsdorfergrund, Germany)
Non-essential Aminoacids	HyClone (IL, USA)
Insulin-transferrin-sodium selenite media supplement	Sigma-Aldrich (MO, USA)

2.1.3 Cell Culture Solutions and Drugs

Complete Growth Medium (DMEM and RPMI)

- 10% FBS
- 1% L-Glutamine
- 1% Penicillin-Streptomycin
- 450 ml Medium

Complete Growth Medium (Ham's F12)

- 10% FBS
- 1% Insulin-transferrin-sodium selenite media supplement
- 1% L-Glutamine
- 1% Penicillin-Streptomycin

L-Glutamine, FBS and Penicillin-Streptomycin was filtered through 0.2 um Millex-FG syringe filters (Merck Millipore, MA, USA) while adding to medium.

Phosphate Buffered Saline (10X PBS)

- 80g Sodium Chloride
- 2g Potassium Chloride
- 2.4g Potassium Phosphate
- 14.4g Sodium Phosphate
- Bring to 1 liter with ddH_2O

10X PBS was firstly diluted to 1X and then autoclaved. 1X PBS was filtered through surfactant-free cellulose membrane serum filter (Thermo Fischer, MA, USA) before use in cell culture.

Freezing Medium

- 10% DMSO
- 90% FBS

2.1.4 Datasets

E-MTAB-2706 RNA-seq dataset were downloaded from ArrayExpress ("http://www.ebi.ac.uk/array express"). E-MTAB-2706 is RNA-seq dataset which

consists of 675 commonly used human cancer cell lines. Dataset contains normalized gene expression read counts for all coding and non-coding genes. There are 144 lung cancer cell lines, 100 lymphoid cancer cell lines and breast, ovary, skin colorectal, pancreas, brain, stomach, head-neck, liver, kidney, sarcomatoid, cervix, urinary bladder, uterus cancer cell lines in decreasing number[82]. In this present study, RNA-seq data of breast, colon and skin cancer cell lines for coding and non-coding genes was used to reveal mechanism which controls CT gene expression. Analyzed cell lines are 52 colon, 70 breast and 49 skin cancer cell lines representing CT-poor, CT-moderate and CT-rich tumors respectively.

In this present study, we have used drug data from CCLE which has pharmacologic profiles for 24 anticancer agents across 504 cell lines. We also used drug sensitivity data from Cancer Genome Project (CGP) database. It includes 265 compounds and 1074 cancer cell lines.

2.2 METHODS

2.2.1 Cell Culture Techniques

Breast cancer cell lines CAL-51, HBL-100, HCC38, MDA-MB-231, MDA-MB-231, MDA-MB-436, BT-20 and SUM149PT were used in this study. CAL-51, HBL-100, MDA-MB-231, MDA-MB-231and MDA-MB-436 cell lines were maintained with complete growth medium DMEM. HCC38 cell line was maintained with complete growth medium RPMI. SUM149PT cell line was maintained with Ham's F12 medium which enriched by 10% FBS, 1% L-glutamine, 1% penicillin-streptomycin and 1% Insulin-transferrin-sodium selenite media supplement. These cell lines were incubated at 37^{o}C in the condition of 5% CO_2. Before seeding, all cell lines were stored in liquid nitrogen. To isolate RNA and perform cytotoxicity experiment with these cell lines, cells were taken from nitrogen tank and thawed immediately at 37C in water bath. Before the cell was completely thawed, pre-warmed complete medium was added to cryotubes to complete thawing. Then, cells were taken into 15 ml falcon tubes and centrifuged for 5 minutes at 1200 rpm to pellet cells. DMSO containing freezing medium was removed by aspirator and pellet was re-suspended in 3 ml of complete growth medium, then added to T-25 flask containing 3 ml complete growth medium. After reaching appropriate confluency, cells were passaged and transferred into 100 mm culture dishes and T-75 flasks. Before passaging, growth medium in the flask was removed by aspirator and cells were washed by sterile 1X PBS. Passaging of cells was done by adding pre-warmed 1 ml Trypsin to detach cells from flask surface. Trypsin added flasks were placed into incubator for 5 minutes until detaching was observed. 4 ml complete growth medium was added to flask to inactivate trypsin and cells were dispersed by pipetting up and down. Then, these cells in media were centrifuged to remove media containing trypsin and dead cells. Pellets then were re-suspended in 4 ml complete growth media and added into flask containing 6 ml media. When the cells were reached confluency and required passaging, washing with PBS, detaching by trypsin, inactivation of trypsin, removal of trypsin by centrifugation and transferring into new flask were performed respectively. When experiments were completed, cells

were cryopreserved. For cryopreservation of cells, media were removed by aspirator and cells washed by 1X PBS once. After removal of PBS by aspirator, cells were detached by trypsinization by using 1 ml Trypsin. By adding 3 ml growth media, trypsin were inactivated and removed by centrifugation. Pellets were then re-suspended in freezing medium containing 90% FBS and 10% DMSO. Cells in freezing medium were then transferred into cryovials. Cryovials containing cells in 1 ml freezing medium were immediately placed into -20^{o}C then transferred into -80^{o}C. For long term storage, stocks were maintained at liquid nitrogen.

2.2.2 Harvesting of Cells for RNA isolation

All cell lines used in this study were adherent cells. For the collection of cells for RNA isolation, growth medium in T -75 flask containing monolayer, 80-90% confluent cells was removed by aspirator and washed with 1X PBS to discard remained medium and dead cells. Then, 1ml TRIzol reagent was added directly to cells in the T-75 flask per 10 cm2 of surface area. Cells were further detached by cell scratcher and homogenized by pipetting up and down. Homogenized samples in TRIzol reagent were put into eppendorf tubes and stored into -80C for later RNA isolation.

2.2.3 Total RNA Isolation with TRIzol

1 ml homogenized samples stored in -80^{o}C were thawed and incubated for 5 minutes at room temperature to allow complete dissociation of the nucleoprotein complex. After adding 0.2 mL of chloroform to homogenized samples, eppendorfs were shaken vigorously by hand for 15 seconds and then, incubated for 3 minutes at room temperature. Samples were centrifuged at 13000 rpm at 4^{o}C for 15 minutes and upper phases containing RNAs were removed by anfling the tubes and pipetting the upper phases out. Upper aqueous phase was taken into new tube and RNA in samples was precipitated by adding 100% 0.5 ml isopropanol to aqueous phase. There was 10 minutes incubation at room temperature for 10 minutes after addition of isopropanol. Incubated samples were then centrifuged at 13000 rpm for 15 minutes at 4^{o}C. Pellets were visible after this step. Then, pellets were washed with 1 ml 75% ethanol after

removal of supernatant. Washing was improved by vortexing tubes. After that, tubes were centrifuged at 10000 rpm for 5 min at 4^{o}C and ethanol was discarded. Pellets were then air dried for 20 minutes. Air dried RNA pellets were then re-suspended in 40 ul of DEPC-treated water. Re-suspended RNA pellets were incubated at 60^{o}C for 15 minutes to dissolve further. RNA samples were stored at -80^{o}C for further experiments.

2.2.4 DNAseI Treatment of Isolated RNA

After phase separation by adding chloroform, upper phases may be contaminated with interphases containing DNA. Elimination of such contaminations was accomplished by DNase treatment using DNA-free kit. RNA samples were firstly diluted to 200 ng/ul. In reaction setup, 4 ul 10X DNase I Buffer and 1 ul rDNase I were added to 40 ul RNA samples. Total reaction volume was incubated at 37^{o}C for 30 minutes in ABI 9700 PCR machine. After incubation, 4.5 ul of pre-mixed DNase inactivation reagent was added into total reaction. Mixing occasionally, reactions were incubated for 2 minutes in room temperatures. With centrifugation for 2 minutes at 13000 rpm, supernatant containing treated RNA was transferred into a fresh eppendorf tube.

2.2.5 RNA Quantification

Isolated RNA from both TRIzol extraction and DNase treated samples were qualified and quantified by NanoDrop One, Merinton Instruments and Qubit RNA BR Assay Kit. With NanoDrop One Instrument, RNA samples were both qualified and quantified applying 2 ul of samples. The instrument has given concentration values by eliminating possible contaminations. Concentrations of RNA samples were also quantified by using Merinton SMA1000 spectrophotometer by applying 2 ul of samples. After DNase treatment, concentrations of RNA samples were quantified by Qubit Fluorometer. To quantify RNA, working solution was firstly prepared by adding 1 ul BR reagent to 199 ul of buffer and then standards were prepared by adding 10 ul of standard 1 and 2 to 190 ul of working solution and RNA samples were prepared by adding 4 ul samples to 196 ul of working solution for every sample. After 2 minutes

incubation, quantities of RNA in samples were measured. The concentration of samples was calculated with following equation:

Concentration of sample= QF value x (200/x)

QF was the value given by Qubit 2.0 Fluorometer and x was the amount of the sample in the reaction setup.

2.2.6 cDNA Synthesis

In synthesis, Thermo Scientific Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Cat# K1641) was used . cDNA synthesis reaction was performed as a series of parallel reactions with DNase-treated RNA samples isolated from all cell lines. 2000 ng total RNA template was used per reaction. 8 ul 5X Reaction Mix and 4 ul Maxima Enzyme Mix, 2000 ng template RNA and nuclease-free water were added up to 40 ul of total reaction volume in a sterile PCR tubes. The tubes were gently mixed and centrifuged briefly. Then, the tubes were incubated at 25^{o}C for 30 minutes followed by 15 minutes at 50^{o}C. cDNA synthesis reaction was terminated by heating at 85^{o}C for 5 minutes. Control reactions were performed with reverse transcriptase minus negative controls which contain every reagent except Maxima enzyme mix and no template control which contain all reagents except any RNA template.

2.2.7 qRT-PCR Analysis for CT Gene Expression

qRT-PCR reactions were performed in triplicates for CT genes and duplicates for housekeeping gene in Roche LightCycler 480 II machine. No template control reactions were also performed for each gene. TaqMan Gene Expression Assays used in this study include 4352934E for GAPDH, Hs03805505_mH for PAGE2 and PAGE2B, Hs02387419_gH for SPANXB, Hs04190522_gH for MAGEA3, Hs01057958_m1 for MAGEA1 and Hs00265824_m1 for NY-ESO-1 genes. The reaction mixture is prepared as indicated in the next page.

2X TaqMan Gene Expression Master Mix	5 ul
TaqMan Primer-probe Mix	0.5 ul
Nuclease-free Water	2.5 ul
cDNA	2 ul
Total volume	10 ul

Thermal cycle conditions were indicated below.

Step		Time	Temperature
UDG Incubation		2 minutes	50°C
AmpliTaq Gold, UP Enzyme Activation		10 minutes	95°C
45 Cycles	Denature	15 seconds	95°C
	Anneal/Extend	1 minute	60°C
Cooling		1 hour	25°C

The relative gene expression values were calculated by using mean of cycle threshold (CT) values of replicates and using $2^{-\Delta\Delta CT}$ calculation, where

$$\Delta\Delta CT = (CT_{Target}\text{-}CT_{GAPDH})_{Sample} - (CT_{Target}\text{-}CT_{GAPDH})_{Reference}$$

2.2.8 Cell Counting

To seed approximately 5000 cells per well into 96-well plate for drug cytotoxicity experiments, haemocytometer was used to count cells in suspension. Cells were firstly detached by trypsinization and re-suspended in 5 ml of complete growth medium. 10 ul from these cell suspensions were taken into an eppendorf tube and mixed well with 10 ul of trypan blue. After cleaning of haemocytometer and coverslip with 70% ethanol gently, 10 ul of cells stained with trypan blue were loaded into each chamber. The cells in 16-squares in each corner of chamber were counted. Average of four 16-squares was calculated by dividing 4, the value was multiplied with dilution factor 2. Multiplying

the final value with 10^4 was resulted in the approximate number of cells per ml of cell suspension.

2.2.9 Drug Treatment and Luminescent Cell Viability Assay

Before drug cytotoxicity experiments, each cell line were cultured in 100 mm culture dishes. After counting with haemocytometer as described previously, 5000 cells/well in 100 ul were seeded into 96 well plates as depicted in Figure 2.1. After a day, cells were treated with different concentrations of Panobinostat, a pan-HDAC inhibitor. Different concentrations of drugs were prepared by using complete growth medium containing 0.1% DMSO. 100 ul of each drug concentrations were loaded quadruplicate for each cell line and incubated for 3 days at 37^{o}C in CO_2 incubator. After 3 days, 96-well plates were taken out and waited for 30 minutes at room temperatures. Then, 30 ul pre-warmed CellTiter-Glo Reagent was added to each well and plates were shaken vigorously as possible for 10 minutes on horizontal shaker. 200 ul of this suspension were transferred to white opaque plate and OD values were measured by using BioTek Synergy HT microplate reader. Measured OD values were used to calculate the cell viability percentage.

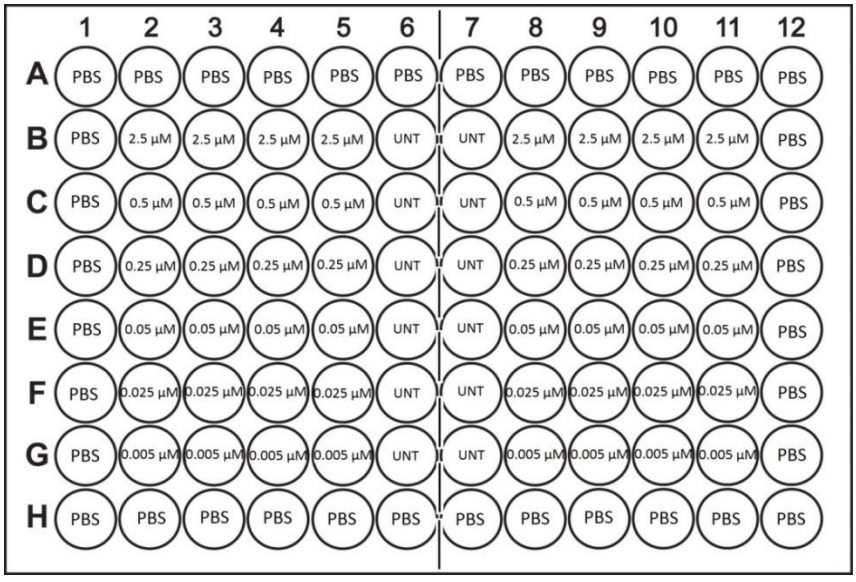


Figure 2.1: Design of drug cytotoxicity experiments.

2.2.10 Calculations of the Half Maximal Inhibitory Concentration (IC50)

Drug cytotoxicity was measured by using CellTiter-Glo Luminescent Cell Viability Assay which is based on quantitation of ATP, an indicator of metabolically active cells. OD values obtained from microplate reader were used to cell viability percentage by using formula given below.

% Cell Viability= (OD value of drug treated well/OD value of 0.1%DMSO treated control) * 100

% cell viability values were then used to construct dose-response curves for each cell lines using GraphPad software. IC50 values for each cell line were analyzed by using GraphPad software.

2.2.11 Software Programs Used in this Study

GraphPad Prism 5.0 was used to construct dose-response curves and analyze half maximal inhibitory concentration. It is also use to draw figures for drug cytotoxicity/CT gene principal component analysis correlation.

2.2.12 Tumors and Adjacent Normal Tissues

All samples used this study were same samples in our previous study [83] obtained from consenting study subjects undergoing surgical tumor resection who signed a written informed consent approved by their respective IRBs.

3. RESULTS

3.1 Revealing Mechanisms which Control CT Gene Expression in Cancer

In the first part of this study, we aimed to identify an experimental model whose study could reveal molecular mechanisms leading to epigenetic changes and subsequent activation of cancer testis genes in cancer. In the first model, we tried to extend our previous observations related to two CT-proximal genes, ALAS2 and CDR1 genes, which showed inverse expression patterns, compared to CT genes in cancer cell lines. We checked this inverse correlation in gene expression in 8 tumors and 8 matched healthy tissues. Then, we performed *in silico* analysis to observe gene expression correlations of CT and CT-proximal genes by using RNA-seq data of breast and colon cancer cell lines. In the second model, we hypothesized that expression levels of CT genes in cancer cell lines can be used to classify cancer cells, whose comparative analysis would help identify mechanisms related to control of CT expression in cancer. For this purpose, cancer cell lines were initially categorized into CT-high, CT-intermediate and CT-low based on their CT gene expression status. In the second step, cell lines were classified based on their epithelial and mesenchymal phenotypes and expression levels of CT genes. With these models, we tried to identify any transcripts that were differentially expressed between these classes and these could help explain mechanisms underlying epigenetic changes and subsequent activation of CT genes.

3.1.1Region Specific Epigenetic Changes Leading to Cancer Testis Gene Expression

Our previous data suggested that those epigenetic changes resulting in expression of CT genes would have to occur within a region in the genome with clear boundaries which would exclude CT proximal non-CT genes. Our group members identified such CT proximal genes using an *in silico* method involving the CGAP database. Genes that

had no expression in cancer cell lines, but had expression in normal healthy tissues were selected. To extend our previous observations related to two such CT proximal genes (ALAS2 and CDR1), and therefore, to validate our earlier findings that were restricted to cell lines and normal tissue, we checked the expression of CT and CT proximal genes in tumor and matched-normal colon tissues (Figure 3.1). In this experiment, we expected to see downregulation of ALAS2 and CDR1 mRNA expression in tumors relatively to normal counterpart. At the same time, we were expecting that expression of CT genes would be upregulated in tumors relative to their normal counterpart. For 3 of tumor and matched normal samples, we observed the expected inverse correlation. However, in the other 5, we could not observe inverse relation between their expression and that of CT genes (PAGE2 and SPANXB). PAGE2 expression, detected in only 2 tumor tissues, was clearly upregulated compared to normal tissues. SPANX-B was detectable in 7 of 8 tumor tissues and was clearly upregulated in 5 tumor tissues. In one case upregulation was present but not obvious, and in one case, tumor proximal tissues showed more SPANX-B expression compared to tumor (#48). ALAS2 expression was clearly detectable in all but two tumors. In 4 tissue pairs, it showed down-regulation in tumors, compared to normal tissues, while it was upregulated in normal tissue compared to tumors in tissues #74 and #126. CDR1 expression was detected in 6 tissues. It was clearly down-regulated in tumors in two tissues, but not in the other 4. However, to our surprise, none but one tumor-pair (#123) showed upregulated CT gene expression concomitant with down-regulated non-CT genes in tumor tissue, compared to its normal counterpart.

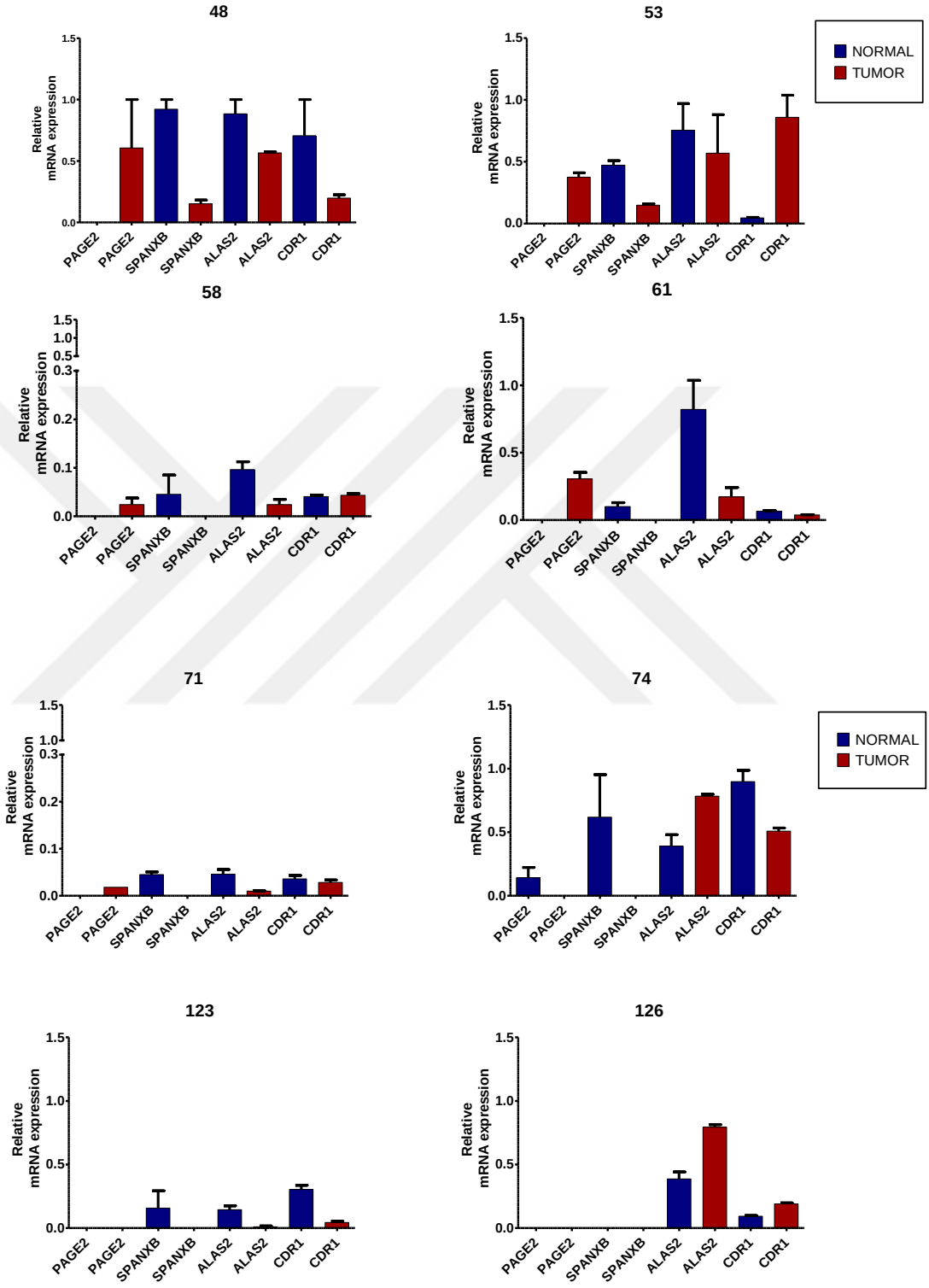

48
53
58
61
71
74
123
126
NORMAL
TUMOR
Relative mRNA expression
PAGE2
PAGE2
SPANXB
SPANXB
ALAS2
ALAS2
CDR1
CDR1

Figure 3.1: mRNA expression levels of PAGE2, SPANXB and proximal genes ALAS2 and CDR1 in tumor and matched normal tissues. Inverse correlation in CT and CT-proximal gene expression was observed only in three tumor-matched normal tissues. mRNA expression of ALAS2, CDR1, PAGE2 and SPANX-B genes was detected with SYBR Green based qRT-PCR. GAPDH gene was used as endogenous control. Values are normalized to the largest expression value obtained for each gene. Blue color indicates healthy tissues; red color indicates its tumor counterpart.

Thus we were unable to show an inverse relationship in expression of CT and CT-proximal genes in tumor and matched-healthy tissues. We hypothesized that this observation could be explained if different cells within the same tissue sample never simultaneously expressed both genes of opposite types. To confirm this, we decided to perform immunohistochemical staining of PAGE2, SPANXB, ALAS2 and CDR1 proteins to show that PAGE2 and SPANXB genes are not expressed in the same cell with ALAS2 and CDR1 genes, respectively. However, as there were no antibodies against ALAS2 and CDR1 proteins, we decided to test *in silico* whether there was an inverse expression pattern by using cell lines. However, inverse relation in CT (PAGE-2, SPANX-B) and CT-proximal (ALAS2, CDR1) gene expression could not be identified among breast and colon cancer cell lines either (Figure 3.2). On the contrary, we even observed a significant positive correlation between SPANXB and CDR1 expression in breast cancer cell lines.

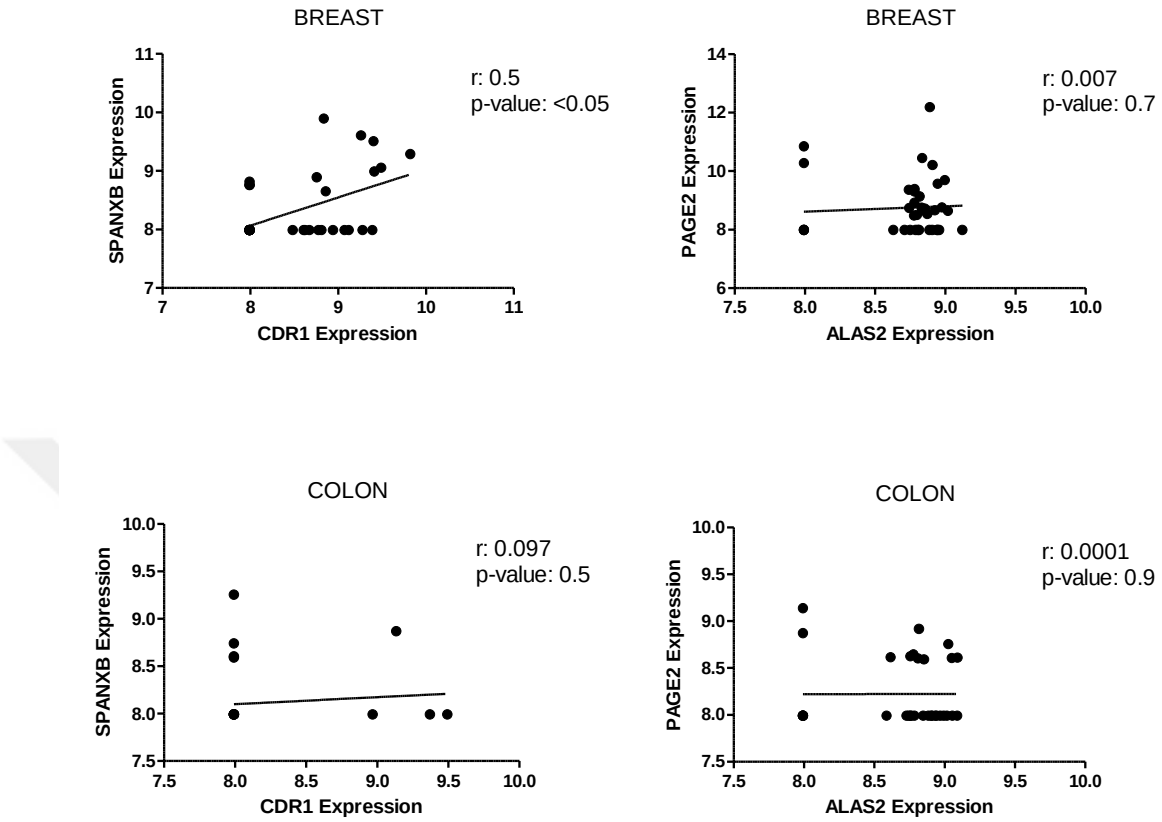


Figure 3.2: Correlation of CT-proximal genes ALAS2/CDR1 expression with PAGE2/SPANXB gene expression in breast and colon cancer cell lines. Inverse correlation in CT and CT-proximal gene expression was not observed among cancer cell lines. Moreover, positive correlation was observed between SPANXB and CDR1 in breast cancer cell lines. Expression data was taken from E-MTAB-2706 dataset [93].

We therefore concluded that identifying region specific epigenetic control mechanisms using the PAGE2/SPANX-B vs. ALAS2/CDR1 model would not be possible. We here could not show inverse correlation in gene expression of CT genes and CT-proximal genes in tumor samples and matched healthy tissue, and also in cancer cell lines while it was shown that ALAS2 and CDR1 genes have expression in normal but they are downregulated in cancer cell lines.

3.1.2 Distribution of CT Gene Expression among Cancer Cell Lines

Since our analysis on CT and CT-proximal gene expression did not support an inverse relation between their expression and that of CT genes and the region specific epigenetic control mechanisms leading CT gene expression, we searched for another experimental model to reveal mechanisms resulting in epigenetic changes and subsequent activation of CT genes in cancer. For this purpose, we hypothesized that if we identified the differential gene expression pattern between CT-expressing and non-expressing cells, then the non-CT genes thus identified would provide clues towards understanding mechanisms that controlled CT gene expression. Therefore, we first identified distribution of CT gene expression across different types of cancer cell lines. A gene list containing 80 CT genes located on X chromosome was used to perform hierarchical clustering to determine the distribution of CT genes among 675 cell lines (Figure 3.3). A heatmap image was created with Cluster 3 and Treeview software. It was observed that some cell lines have low levels of CT gene expression while others have higher levels of CT gene expression with coordinate expression. It was also seen that CT gene expression was very heterogeneous among different types of cell lines; each type has both CT-low and CT-high cell lines. We next classified cancer cell lines based on their expression levels of CT genes.

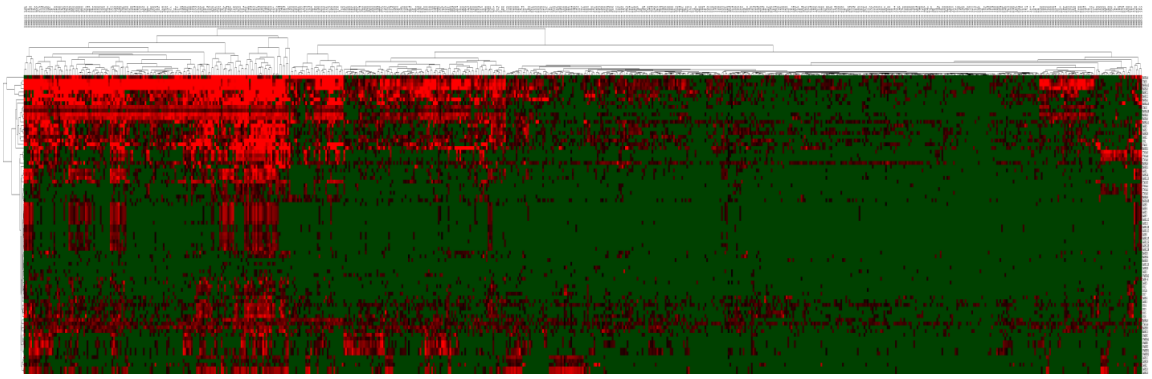


Figure 3.3: Distribution of CT gene expression among cancer cell lines. Distribution of CT gene expression is highly heterogeneous among cancer cell lines similar to intratumoral heterogeneity of CT gene expression. Normalized gene expression data was used to perform hierarchical clustering by using 80 cancer-testis genes. Columns are 675

commonly used cell lines in E-MTAB-2706 dataset. Red represents maximum expression value and green represents minimum expression value, while black is intermediate. Most of cells have distinct expression pattern, and CT genes are coordinately expressed in these cell lines.

3.1.3 Subgrouping Breast, Skin and Colon Cancer Cell Lines based on CT Gene Expression

We first analyzed CT gene expression data of different cancer types. We observed that CT gene expression was highly variable in Breast, Skin and Colon cancer cell lines among different types of cell lines. Therefore, these types of cancer cell lines were chosen for further evaluation by categorizing the cell lines according to their CT gene expression levels. Expression values of CT genes were used for principle component analysis. First principle component was used to classify cell lines into three groups, CT-High, CT-Int, and CT –Low. Cut-off values used to subgroup cell lines was same for all breast, skin and cancer cell lines (Figure 3.4, 3.5, 3.6).

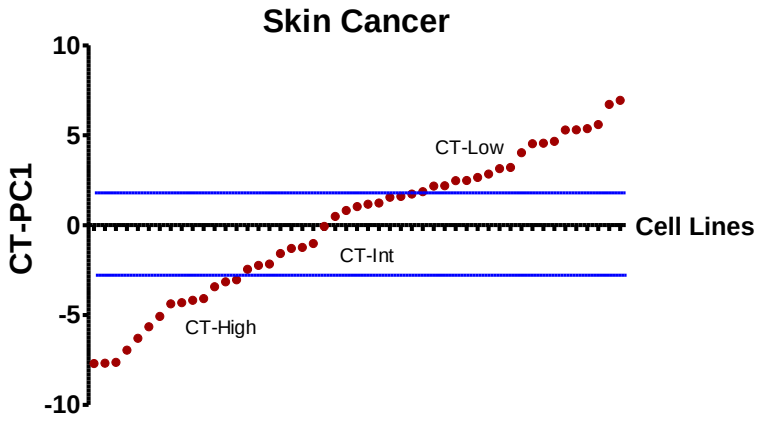


Figure 3.4: Categorizing skin cancer cell lines into CT-High, CT-Int and CT-Low groups. 49 skin cancer cell lines are classified into three groups, CT-High, CT-Int and CT-Low expressors by using first principal component values of 80 CT genes. CT-PC1 values are the first principle component values which are calculated with principle component analysis. Lower PC1 values indicate high levels of CT gene expression in cell lines while cell lines which have low levels of CT gene expression have lower PC1 value. Cut-offs to divide cells into groups was arbitrarily defined to compose equal sample size in each.

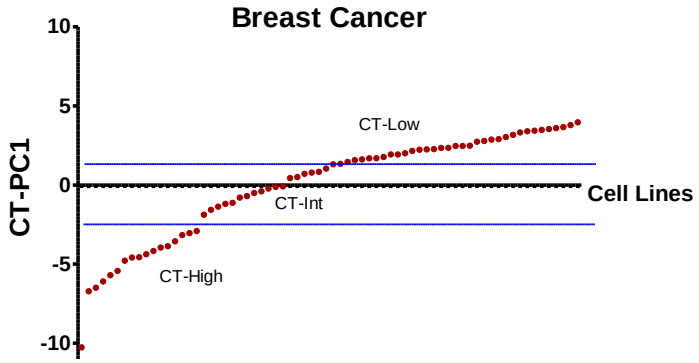


Figure 3.5: Categorizing breast cancer cell lines into CT-High, CT-Int and CT-Low groups. 70 breast cancer cell lines are classified into three groups, CT-High, CT-Int and CT-Low expressors by using first principal component values of 80 CT genes. CT-PC1 values are the first principle component values which are calculated with principle component analysis. Lower PC1 values indicate high levels of CT gene expression in cell lines while cell lines which have low levels of CT gene expression have lower PC1 value. Cut-offs to divide cells into groups was arbitrarily defined to compose equal sample size in each.

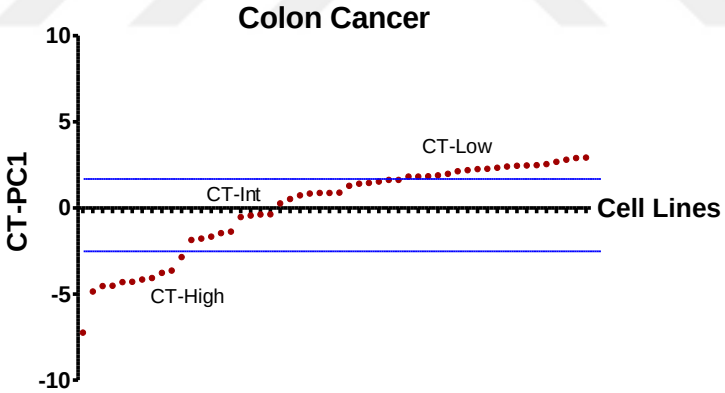


Figure 3.6: Categorizing colon cancer cell lines into CT-High, CT-Int and CT-Low groups. 52 colon cancer cell lines are classified into three groups, CT-High, CT-Int and CT-Low expressers by using first principal component values of 80 CT genes. CT-PC1 values are the first principle component values which are calculated with principle component analysis. Lower PC1 values indicate high levels of CT gene expression in cell lines while cell lines which have low levels of CT gene expression have lower PC1 value. Cut-offs to divide cells into groups was arbitrarily defined to compose equal sample size in each.

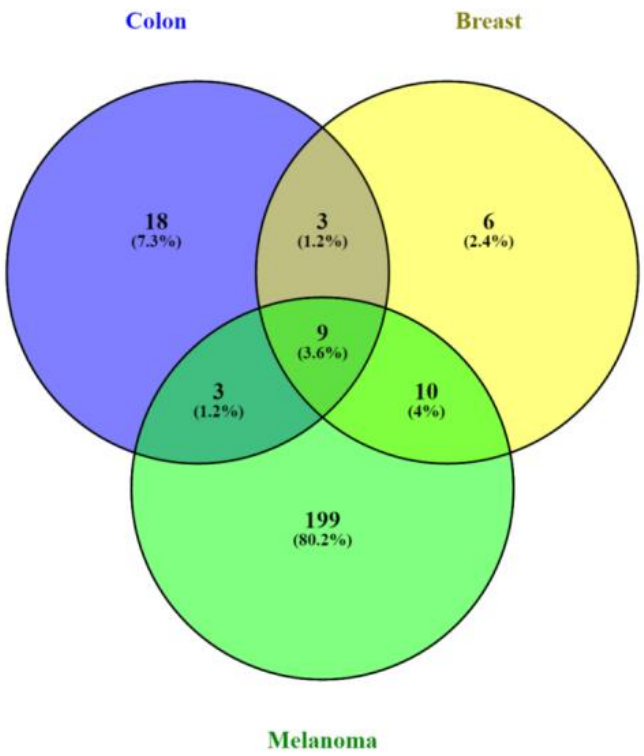
3.1.4 Identification of Differentially Expressed Transcripts (DET) between CT based Subgroups

In order to determine non-CT genes which may control the expression of CT genes in cancer cell lines, we found the differentially expressed transcripts between CT-Low and CT-High groups in each cancer cell type by t-test (Appendix Table 1, 2, 3,). Then, we applied Benjamini-Hochberg correction to p-values of significant genes. Two hundred and twenty one genes and non-coding RNAs were identified as being differentially expressed in skin cancer cell lines when CT-Low and CT-High groups of skin cancer cell lines. Twenty eight and thirty three genes and some non-coding RNAs were identified as being differentially expressed in breast and colon cancer cell lines, respectively. As expected, most of the DETs were CT genes since subgroups were defined based on CT gene expression levels. There were also CT genes which are not included in the gene list used for principal component analysis. It confirmed coordinated expression pattern of cancer testis genes. Furthermore, we hypothesized that common non-CT genes between different cancer types could suggest a common mechanism leading to CT gene expression. Therefore we checked common genes between different cancer cell lines.

3.1.5 Comparison of DETs among Different Cancer Types

To determine whether there are common genes and mechanisms which control the CT gene expression in different cancer types, we decided to compare significant genes in all types of cancer which are found by analyzing differentially genes between CT-High and CT-Low groups. Nine genes were found to be common between colon, breast and skin cancer cell lines. However, all of these genes were CT genes; we could not determine genes which may be the part of some gene clusters controlling CT gene expression. This finding suggested that each cancer types may have specific mechanisms leading to epigenetic changes and subsequent activation of cancer testis

genes. We therefore hypothesized that non-overlapping and distinct mechanisms could be involved in the re-activation of CT genes in different tumors.



Common Genes
MAGEA6
MAGEA2
CSAG2
MAGEA3
CSAG1
GABRA3
MAGEA12
PAGE2B
MAGEA2B

Figure 3.7: Nine differentially expressed genes identified in colon, breast and skin cancer cell lines were common. Twenty genes were common between colon and breast cancer cell lines, twenty genes were common between colon and skin cancer cell lines and nineteen genes were common between skin and breast cancer cell lines. Nine of these genes were common between all three cancer types. These common genes can be seen in the table, right.

3.1.6 Identification of Differentially Expressed Non-coding Transcripts between CT based Subgroups

We also checked the expression pattern of non-coding transcripts involved in dataset, in order to determine non-coding genes which could help explain mechanisms controlling the expression of CT genes in cancer cell lines, we found the differentially expressed non-coding genes between CT-Low and CT-High groups in each cancer cell type by t-test (Appendix Table 4, 5, 6). We identified 3 differentially expressed non-coding RNAs in colon cancer cell lines, 19 in breast and 17 transcripts in skin cancer cell lines. All differentially expressed non-coding transcripts were upregulated in CT-High group of cancer cell lines. It suggested that expression of these non-coding transcripts can be consequence of similar epigenetic changes rather than being a cause of CT gene re-activation in cancer. Also, we found that none of differentially expressed non-coding genes identified in colon, breast and skin cancer cell lines were common (Figure 3.8).

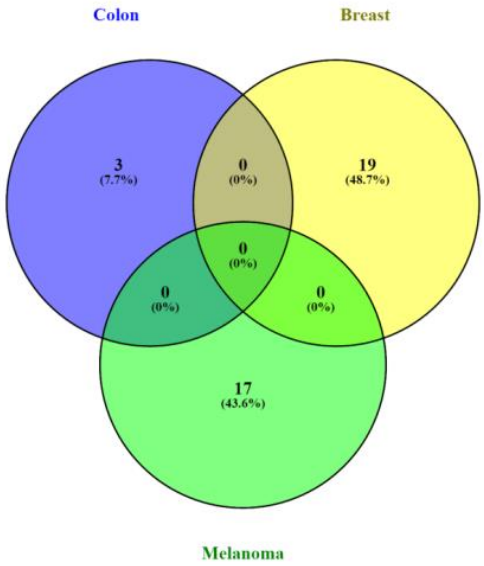


Figure 3.8: Common non-coding transcripts between colon, breast and skin cancer cell lines. None of differentially expressed non-coding genes identified in colon, breast and skin cancer cell lines were common.

In this approach, classifying cancer cell lines based on CT gene expression levels did not help explain mechanisms underlying CT gene activation in cancer as all genes that were identified were either CT genes themselves, or were upregulated, suggesting they followed the same induction mechanisms that lead to CT gene expression. Also, differential expression analysis between CT-High and CT-Low groups did not result in common genes between different cancer types. This suggested that non-overlapping and distinct mechanisms could be involved in the activation of cancer testis genes in different tumors. As our earlier work suggested a relationship between epithelial to mesenchymal transition and CT gene expression we asked if an EMT based classification could help elucidate mechanisms regulating epigenetic changes and subsequent activation of CT genes in cancer.

3.2 Relationship between CT Gene Expression and Epithelial/ Mesenchymal Phenotype

Our recent studies let us to hypothesize that there is a window during EMT in which CT genes are expressed [36]. In this hypothesis, CT gene expression is supposed to be low in highly epithelial and mesenchymal phenotype (Figure 3.9). This was based on the observation that as colon cancer cells differentiate in vitro, they upregulate CT genes [36]. However, as normal cells are CT negative, and since under normal conditions, CT genes are expressed by commited stem cells, the picture needs to be more complex. To elucidate this, and to generate a new hypothesis by which we could identify mechanisms leading to CT gene expression. We aimed to find out expression status of CT genes in cells with the knowledge of their epithelial and mesenchymal phenotype, and subsequently perform a comparison analysis. In skin cancer cell lines, most of cells are homogeneous in phenotype, they are highly mesenchymal and distribution of CT-Highness and CT-Lowness was very heterogeneous (Figure 3.10). We observed that CT gene expression was highly variable in breast cancer cells, concentrating in more epithelial and more mesenchymal cells (Figure 3.10). In colon cancer cell lines, similar distribution was observed and CT-High cells have mesenchymal or epithelial phenotype (Figure 3.11). We realized that our previous

hypothesis suggesting a window during EMT in which CT genes are expressed has to be changed based on new findings (Figure 3.13). A better model would be that CT genes are expressed when the cell is mesenchymal, is downregulated as the cell differentiates into a more epithelial phenotype, until a threshold is reached after which upregulation is observed. This model also helps explain literature that relates CT gene expression in both a stem-cell like phenotype and more epithelial phenotype. However, if the cell could differentiate back into a normal phenotype, we would possibly expect CT gene expression to again show downregulation. But in cancer, this final explanation might hold true.

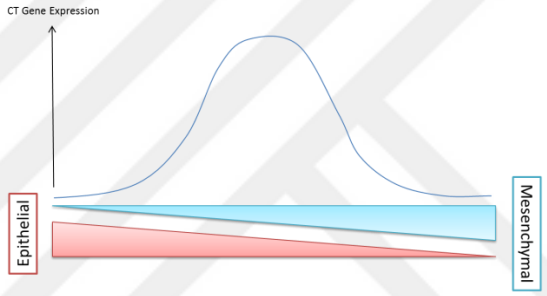


Figure 3.9: Suggested expression patterns CT genes during EMT based on observations in our previous studies.

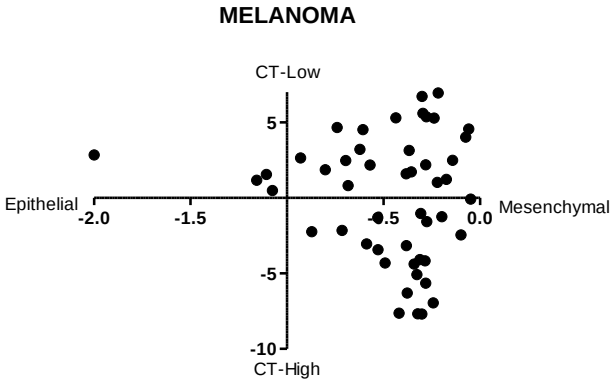


Figure 3. 10: Distribution of skin cancer cells in epithelial/mesenchymal phenotype and CT-expression based classification. Skin cancer cell lines are highly mesenchymal and cells with CT-High or CT-Low expression are heterogeneously distributed. X-axis shows the EMT score of cells while Y-axis shows CT-first principal component values of cells. EMT score is calculated by CDH1-VIM expression based algorithm. First

principal component values for CT gene expression were calculated with R based code. Right-to-left, epithelialness of cells increases. Top-to-bottom, CT gene expression in cells increases.

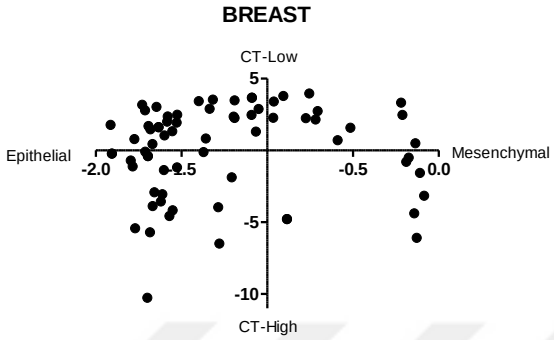


Figure 3.11: Distribution of breast cancer cells in epithelial/mesenchymal phenotype and CT-expression based classification. In breast cancer cell lines, CT-High expressor cells are highly mesenchymal or highly epithelial while mid-phenotype cells have low levels of CT gene expression. X-axis shows the EMT score of cells while Y-axis shows CT-first principal component values of cells. EMT score is calculated by CDH1-VIM expression based algorithm. First principal component values for CT gene expression were calculated with R based code. Right-to-left, epithelialness of cells increases. Top-to-bottom, CT gene expression in cells increases.

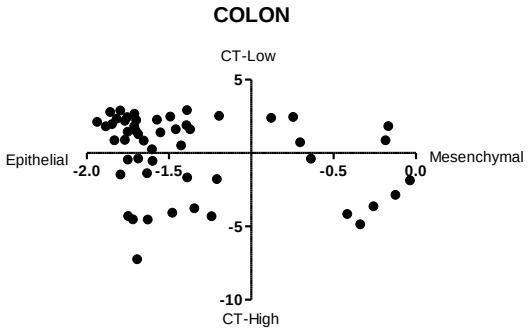


Figure 3.12: Distribution of colon cancer cells in epithelial/mesenchymal phenotype and CT-expression based classification. In colon cancer cell lines, CT-High expressor cells are highly mesenchymal or highly epithelial while mid-phenotype cells have low levels of CT gene expression. X-axis shows the EMT score of cells while Y-axis shows CT-first principal component values of cells. EMT score is calculated by CDH1-VIM expression based algorithm. First principal component values for CT gene expression were calculated with R based code. Right-to-left, epithelialness of cells increases. Top-to-bottom, CT gene expression in cells increases.

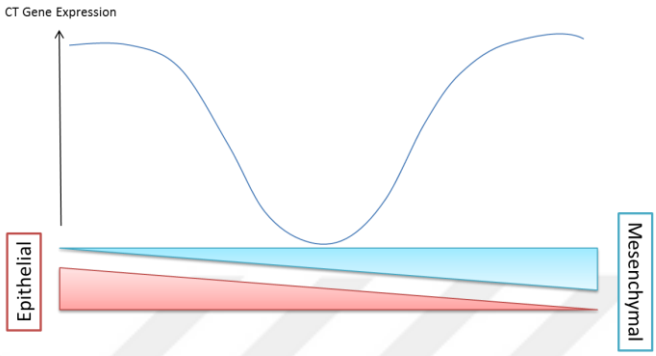


Figure 3.13: New model for CT gene expression in EMT. Suggested expression pattern of CT genes in EMT from previous hypothesis has been changed with new findings. X-axis represents EMT; y-axis represents level of CT gene expression. CT genes are expressed in a window during EMT (left). CT gene expression is high in more mesenchymal or more epithelial phenotype (right).

3.2.1 Subgrouping based on E/M phenotype and CT Gene Expression

We therefore, thought that if there are different mechanisms controlling CT gene expression in different cells, as determined by their EMT phenotypes, lumping all cell lines to find differentially expressed transcripts could mislead us, especially if CT high mesenchymal cells use mechanisms very different from CT high epithelial cells. This is supported by the high numbers of differentially expressed transcripts identified in melanoma with previous approach since they are very homogeneous (mostly mesenchymal) in phenotype. However, the presence of both epithelial as well as mesenchymal cells in breast and colon cancer cell lines may cause elimination of significant genes while comparing CT-High and CT-Low groups. According to this hypothesis, we firstly divided cell lines into epithelial and mesenchymal subtype, and then we determined CT-High and CT-Low cell lines analyzing first principal component with CT gene list. Breast and colon cancer cell lines were used for this analysis, since skin cancer cell lines are highly mesenchymal phenotype. Now, we have

four different subgroups; CT-High Epithelial, CT-Low Epithelial, CT-High Mesenchymal and CT-Low Mesenchymal.

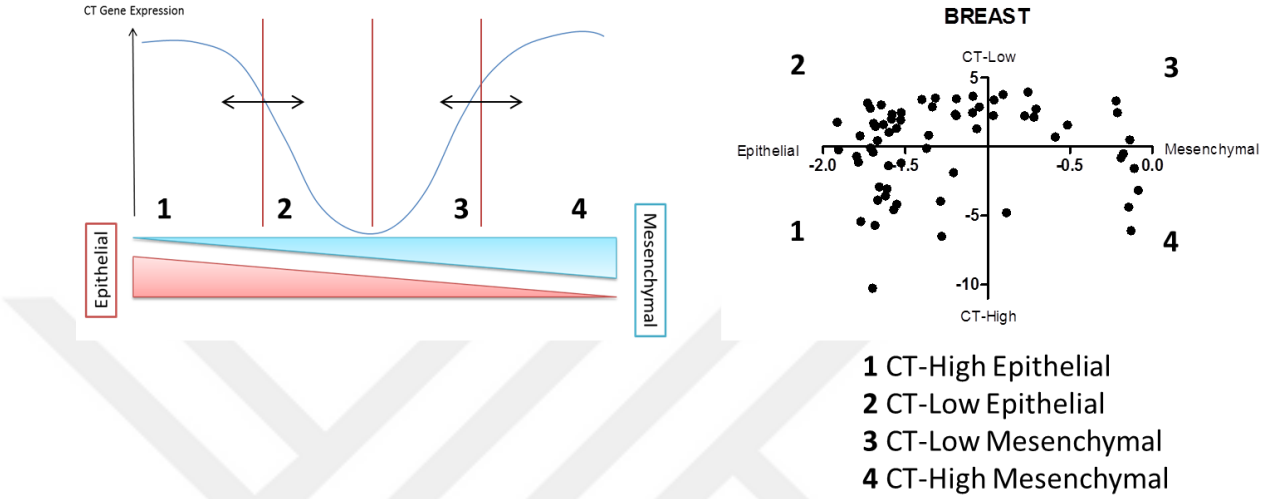
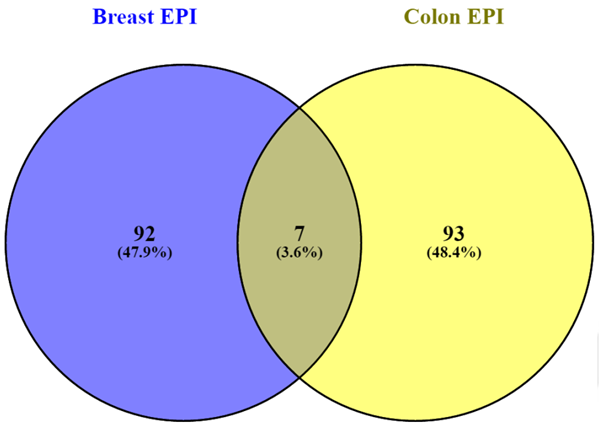


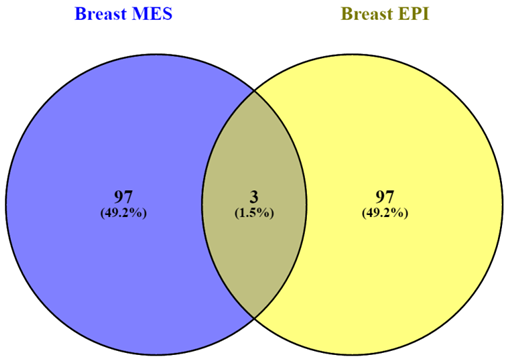
Figure 3.14: Categorizing cell lines based on EMT status and CT gene expression levels. X-axis represents EMT; y-axis represents level of CT gene expression. CT gene expression is high in more mesenchymal or more epithelial phenotype (left). In the right, subgroups were shown in breast cancer cell lines.

3.2.2 DET Analysis with New Categorization

Considering epithelial/mesenchymal phenotype of cell lines, we found out the differentially expressed genes between CT-High and CT-Low cells to reveal the genes which may control CT gene expression in cancer by explaining the epigenetic mechanism behind it. We determined top 100 genes in differential expression analysis by comparing CT-High and CT-Low groups (Appendix Table 7, 8, 9, 10, 11). Then, we compared differentially expressed transcripts between epithelial breast and colon cancer cell lines. There were 7 common genes but all of them were CT genes. When we compare differentially expressed transcripts in epithelial and mesenchymal phenotypes of breast cancer cell lines, common genes were CT genes again (Figure 3.15). In colon cancer cell lines, we observed similar results (Figure 3.16). These findings suggested that there could still be heterogeneity among tumors which would have to be defined which led to distinct mechanisms in different cancer types.



Common Genes
CSAG2
MAGEA3
MAGEA2
CSAG1
MAGEA12
MAGEA6
MAGEA2B



Common Genes
MAGEA3
MAGEA2
MAGEA2B

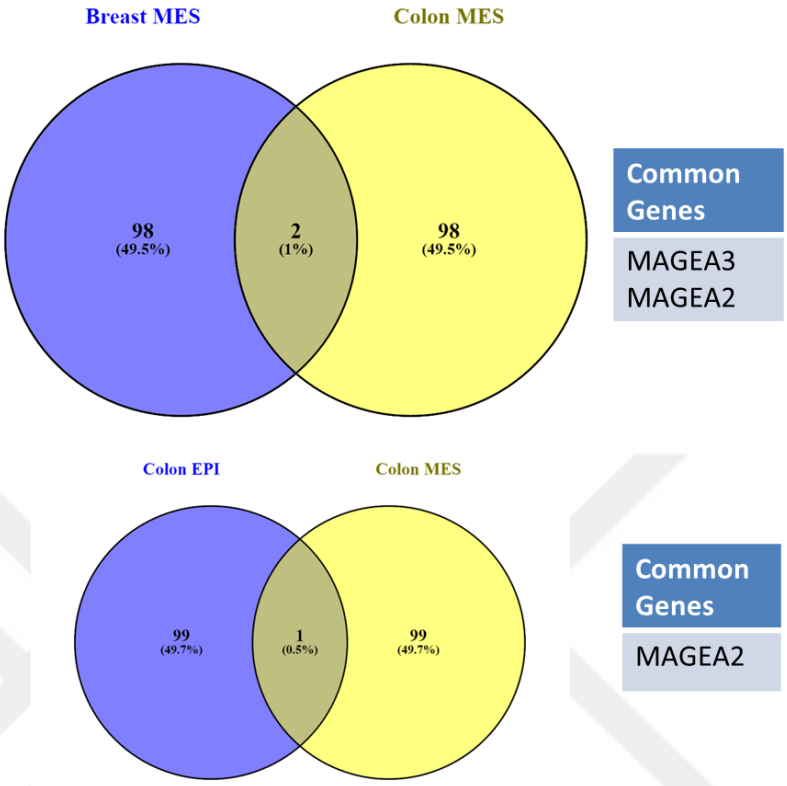


Figure 3.15: Common genes between differentially expressed genes identified by new categorization. At the top of figure, differentially expressed genes in CT-High and CT-Low groups of epithelial breast and colon cancer cell lines were compared to find common genes. There were 7 common genes. Secondly, differentially expressed genes in CT-High and CT-Low groups of epithelial and mesenchymal breast cancer cell lines were compared to find common genes. There were 3 common genes. Thirdly, differentially expressed genes in CT-High and CT-Low groups of mesenchymal breast and colon cancer cell lines were compared to find common genes. There were 2 common genes. Finally, differentially expressed genes in CT-High and CT-Low groups of epithelial and mesenchymal colon cancer cell lines were compared to find common genes. One common gene was MAGEA2, a CT gene.

3.2.3 GSEA with CT-High and CT-Low Subgroups Classified by Epithelial/Mesenchymal Phenotype

In order to provide comprehensive information on mechanisms leading to activation of CT genes, here we employed expression data of CT-High and CT -Low cancer cell lines and applied gene set enrichment analysis to compare gene expression profiles between them (Appendix Table 12, 13). Differential genes and differentially activated signaling pathways were discovered when cancer cell lines were grouped by their EMT status (Table 3.1, 3.2, 3.3, 3.4, 3.5, 3.6).

Table 3.1: Enriched gene sets in CT-High epithelial breast cancer cell lines.

Gene set	NES	FWER p-val
UP Genes when BMI1/MEL18 DOWN	2.188654	<0.001
Genes regulated by NFKB	2.079312	<0.001
UP Genes when BMI1 DOWN	2.034547	<0.001
UP Genes when P53 DOWN	1.992454	0.001
UP Genes when MEL18 DOWN	1.97981	0.001
UP Genes when PTEN DOWN	1.971668	0.001
DOWN Genes during ESC differentiation	1.968822	0.001
UP Genes with active oncogenic KRAS	1.930892	0.002
UP Genes with active RAF1 gene	1.929605	0.002
Genes regulated by NFKB	1.839982	0.009
UP Genes when NFE2L2 Knockout	1.819604	0.011
Genes regulated by Hippo Pathway	1.757583	0.03
UP Genes with active EGFR gene	1.68434	0.087

Table 3.2: Enriched gene sets in CT-Low epithelial breast cancer cell lines.

Gene set	NES	FWER p-val
DOWN Genes with active RAF1 gene	-2.290925	<0.001
DOWN Genes with active EGFR gene	-1.7846069	0.017
DOWN Genes with active MAP2K1 gene	-1.7042557	0.035

Table 3.3: Enriched gene sets in CT-High mesenchymal breast cancer cell lines.

Gene Set	NES	FWER p-val
Genes regulated by Hippo Pathway	2.333679	<0.001
UP Genes with active EGFR gene	2.093535	<0.001
UP Genes when BMI1/MEL18 DOWN	2.032252	<0.001
DOWN Genes during ESC differentiation	1.936375	0.001
UP Genes when MEL18 DOWN	1.931084	0.001
UP Genes when RPS14 DOWN	1.929102	0.001
UP Genes in Astroglial	1.919633	0.002
UP Genes when RB1/RBL1 Knockout	1.845556	0.008
UP Genes when STK33 DOWN	1.826537	0.008
UP Genes when HOXA9 DOWN	1.790701	0.015
UP Genes with active oncogenic KRAS	1.759516	0.031
UP Genes when EIF4GI DOWN	1.758532	0.031
UP Genes when PTEN DOWN	1.754167	0.032
Genes regulated by NFKB	1.750758	0.032
UP Genes when AKT1 DOWN	1.735756	0.042
UP Genes when BMI1 DOWN	1.731516	0.046

Table 3.4: Enriched gene sets in CT-Low mesenchymal breast cancer cell lines.

Gene Set	NES	FWER p-val
Genes with KRAS Dependency	-2.22262	<0.001

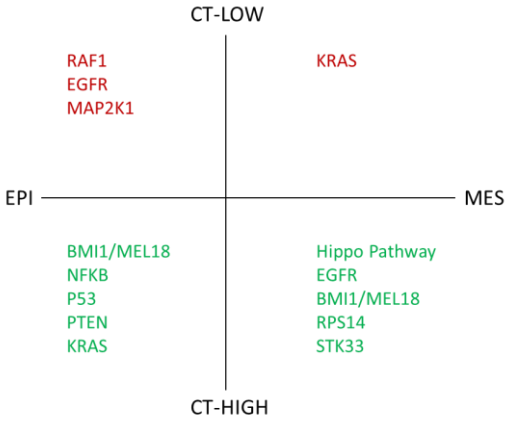


Figure 3.16: Summary of gene set enrichments in different phenotypes with different CT expression levels in breast cancer cell lines.

Comparison of enriched gene sets between different phenotypes in CT-High and CT-Low groups suggest that different proteins and signaling pathways may have role in regulating CT gene expression in different subgroups of new approach. Hippo pathway related gene set was highly enriched in CT-High group of mesenchymal breast cancer cell lines. Upregulated genes when transcriptional repressor proteins BMI1 and MEL18 were knockdown were highly enriched in CT-High epithelial breast cancer cell lines while enrichment of this gene set was in third place in CT-High mesenchymal breast cancer cell lines suggesting that in distinct phenotypes, CT gene expression levels can be regulated by different pathways or proteins in coordinated fashion. Hippo pathway would compensate regulatory function of BMI1 and MEL18 proteins on CT gene expression.

Table 3.5: Enriched gene sets in CT-High epithelial colon cancer cell lines.

Gene Set	NES	FWER p-val
UP Genes with active oncogenic KRAS	1.7210668	0.019
UP Genes with LEF1 UP	1.6946179	0.031

Table 3.6: Enriched gene sets in CT-Low epithelial colon cancer cell lines.

Gene Set	NES	FWER p-val
DOWN Genes with LEF1 UP	-2.1312501	<0.001
Genes regulated by NFKB	-2.1145008	<0.001
UP Genes with active CTNNB1	-1.9069889	0.001
UP Genes with active RAF1	-1.8498989	0.004
DOWN Genes when EIF4E overexpression	-1.8379946	0.005
DOWN Genes when TBK1 knockdown	-1.7889806	0.009
DOWN Genes with active MAP2K1	-1.7755474	0.009
UP Genes when E2F3 overexpression	-1.7652453	0.009
UP Genes when EIF4GI knockdown	-1.7183368	0.017
DOWN Genes in estrogen-independent growth	-1.6195152	0.044
DOWN Genes when RB1 and RBL1 knockout	-1.6057601	0.048

Table 3.7: Enriched gene sets in CT-High mesenchymal colon cancer cell lines

Gene Set	NES	FWER p-val
UP Genes with oncogenic KRAS	2.696188	<0.001
UP Genes in neurons	1.991552	<0.001
DOWN Genes with mutated P53	1.948741	0.001
DOWN Genes when MEL18 DOWN	1.844684	0.004
UP Genes when PTEN UP	1.804608	0.007
UP Genes when E2F3 UP	1.730074	0.023

Table 3.8: Enriched gene sets in CT-Low mesenchymal colon cancer cell lines

Gene Set	NES	FWER p-val
DOWN Genes during ESC differentiation	-2.50182	<0.001
UP Genes in Astroglia cells	-2.17781	<0.001
UP Genes when BMI1 DOWN	-2.13674	<0.001
UP Genes when P53 mutated	-2.118	<0.001
UP Genes when MEL18 DOWN	-2.03904	<0.001
UP Genes with LEF1 UP	-2.03804	<0.001
UP Genes with active EGFR	-2.01646	<0.001
UP Genes with active EGFR	-1.97349	<0.001
Genes regulated by NFKB	-1.88745	0.002
UP Genes by TGFB1	-1.86075	0.003
Genes regulated by NFKB	-1.84934	0.004
UP Genes with oncogenic KRAS	-1.83665	0.006
DOWN Genes with MYC oxerexpression	-1.82518	0.006
DOWN Genes in early stages of differentiation	-1.82079	0.006
UP Genes when BMI1/MEL18 DOWN	-1.81835	0.006
Genes with KRAS Dependency	-1.81004	0.006
UP Genes with active CTNNB1	-1.77774	0.011
DOWN Genes when RB1 DOWN	-1.7063	0.026
UP Genes with IL2 UP	-1.69594	0.028

3.3 Potential Clinical Value of CT Gene Expression

While studying distribution of CT gene expression with latest experimental model, high levels of CT gene expression was observed in cells with more epithelial and more mesenchymal phenotypes. We first asked whether CT gene expression could be biomarker for predicting chemosensitivity. To study potential clinical relevance of CT gene expression in cancer, we decided to study on breast cancer cell lines. First, we searched for any correlation in CT gene expression and drug sensitivity by using whole cell lines without categorizing into breast cancer intrinsic subtypes. However, we could not obtain significant correlation by using expression data for whole cell lines (Table 3.7). Then, we focused on subtype specific expression pattern of CT genes (Figure 3.17). In basal Luminal and Basal A subtypes, we could not define any good correlation with CT gene expression and drug sensitivity. Then, we found that panobinostat sensitivity of Basal B breast cancer cell lines has good correlation with CT expression levels in these cell lines (Table 3.7). We determined a cut-off for pearson r correlation between CT gene expression and drug response data for each anti-cancer agents in CCLE drug database. 0.7 was our cut-off which defines good correlation. Panobinostat sensitivity was highly correlated with increasing levels of CT gene expression.

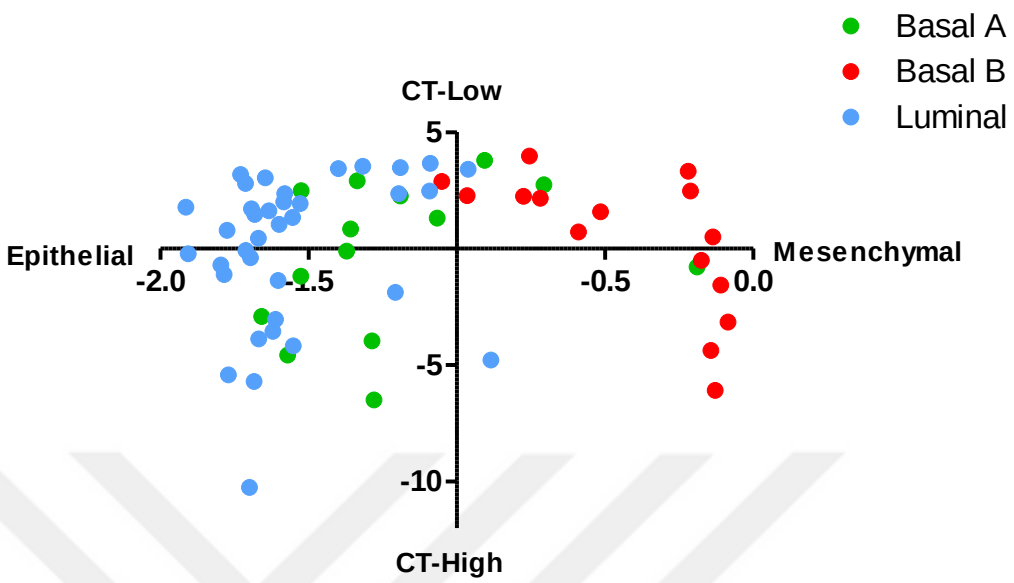


Figure 3.17: CT gene expression within intrinsic subtypes of breast cancer cell lines. Basal B subtype of breast cancer cell lines have mostly mesenchymal phenotype in contrast to Luminal and Basal A subtypes which are more epithelial. X-axis shows the EMT score of cells while Y-axis shows CT-first principal component values of cells. EMT score is calculated by CDH1-VIM expression based algorithm. First principal component values for CT gene expression were calculated with R based code. Right-to-left, epithelialness of cells increases. Top-to-bottom, CT gene expression in cells increases. Green dots, Basal A subtype; red dots, Basal B subtype; Blue dots, Luminal subtype.

Table 3.9: Correlations of CT gene expression and drug response in breast cancer cell lines.

All Cells		Luminal		Basal A		Basal B	
Drugs	pearson r	Drugs	pearson r	Drugs	pearson r	Drugs	pearson r
Paclitaxel	-0.52454	Topotecan	-0.63082	Nutlin-3	-0.51002	Panobinostat	-0.77593
Topotecan	-0.48734	Paclitaxel	-0.59623	LBW242	-0.43785	TKI258	-0.55811
RAF265	-0.43372	RAF265	-0.545	PD-0325901	-0.43062	PF2341066	-0.52305
Erlotinib	-0.38274	LBW242	-0.50516	Panobinostat	-0.41853	PD-0325901	-0.47637
LBW242	-0.34404	ZD-6474	-0.4842	Erlotinib	-0.40027	Nutlin-3	-0.3601
ZD-6474	-0.33558	Nilotinib	-0.41946	RAF265	-0.38815	TAE684	-0.34532
Panobinostat	-0.32506	PD-0332991	-0.34586	PLX4720	-0.38401	L-685458	-0.28112
L-685458	-0.27214	L-685458	-0.34098	Nilotinib	-0.37046	PD-0332991	-0.26763
Nilotinib	-0.22639	TKI258	-0.29129	TAE684	-0.34839	Nilotinib	-0.20303
Nutlin-3	-0.20967	PHA-665752	-0.28048	Sorafenib	-0.34536	Erlotinib	-0.10033
PHA-665752	-0.18294	Panobinostat	-0.27096	TKI258	-0.33283	Lapatinib	-0.04635
TKI258	-0.15631	Sorafenib	-0.24312	ZD-6474	-0.30663	PHA-665752	-0.01697
PD-0332991	-0.15356	PLX4720	-0.2399	Topotecan	-0.22295	Paclitaxel	0.040104
Lapatinib	-0.14162	Nutlin-3	-0.20606	Lapatinib	-0.21113	Topotecan	0.179652
Sorafenib	-0.09634	Lapatinib	-0.15379	PD-0332991	-0.17813	RAF265	0.179937
PD-0325901	-0.08087	Erlotinib	-0.03853	Paclitaxel	-0.14859	ZD-6474	0.239173
PLX4720	-0.07082	PF2341066	0.184837	L-685458	-0.08644	Sorafenib	0.313684
TAE684	-0.05415	TAE684	0.234851	PF2341066	-0.03407	PLX4720	0.525192
PF2341066	0.156941	PD-0325901	0.441687	PHA-665752	-0.01635	LBW242	0.542801

3.3.1 Panobinostat sensitivity Correlated with CT Gene Expression in Basal B Subtype

Using CCLE and CGP drug databases and their expression data, we searched correlation between CT gene expression and Panobinostat sensitivity in Basal B subtype breast cancer cell lines (Figure 3.18). We observed that CT gene expression were high in cells which are more sensitive to Panobinostat treatment. Cells with lower CT gene expression seem to be relatively resistant to Panobinostat. We also observed that CT high cell lines are sensitive to Dacinostat, another HDAC inhibitor. However, this correlation was not observed in other HDAC inhibitor Vorinostat (Figure 3.19). Sensitivity to other anti-cancer agents did not show good correlations with CT gene expression levels in cell lines (Figure 3.20). To validate this correlation, we studied with 7 Basal B and 1 Basal A cell lines by treating Panobinostat.

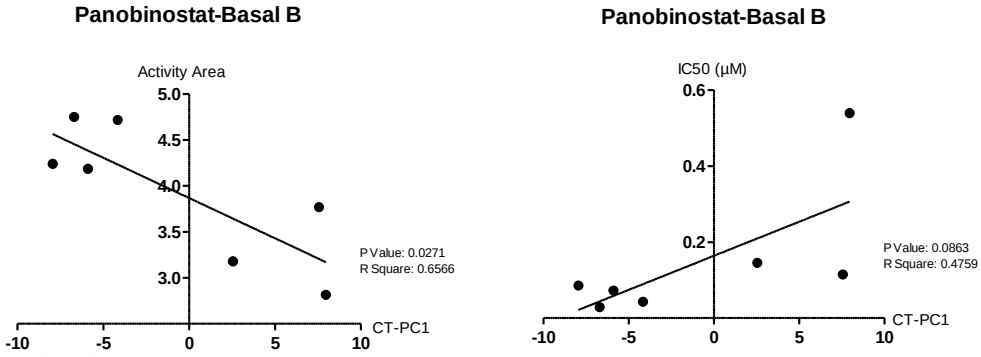


Figure 3.18: Panobinostat sensitivity correlation of CT-PC1 in Basal B cells. CT-High Basal B subtype-breast cancer cell lines are sensitive to Panobinostat, a pan-HDAC inhibitor. CCLE database contains pharmacologic profiles for 24 anticancer drugs across 504 cell lines. 28 of these cell lines are derived from breast, 7 of 28 are Basal B subtype-breast cancer cell lines. Pearson correlation analysis was performed by using drug response values (Activity area and IC50) and CT gene expression score calculated by first principal component analysis.

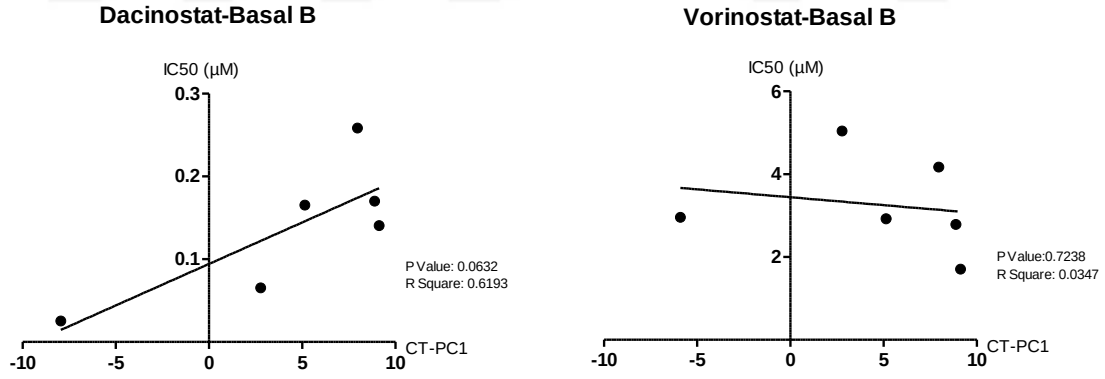


Figure 3.19: Dacinostat and Vorinostat sensitivity correlation of CT-PC1 in Basal B cells. Similar correlation was observed with Dacinostat but not with Varinostat, HDAC inhibitors. Drug response values for CAL-120, Hs 578T, MDA-MB-157, MDA-MB-231, BT-549, HCC38, CAL51, HCC1395 Basal B cell lines were taken from CGP database. 42 of these cell lines are derived from breast, 8 of 42 are Basal B subtype-breast cancer cell lines. Pearson correlation analysis was performed by using drug response values (IC50) and CT gene expression score calculated by first principal component analysis.

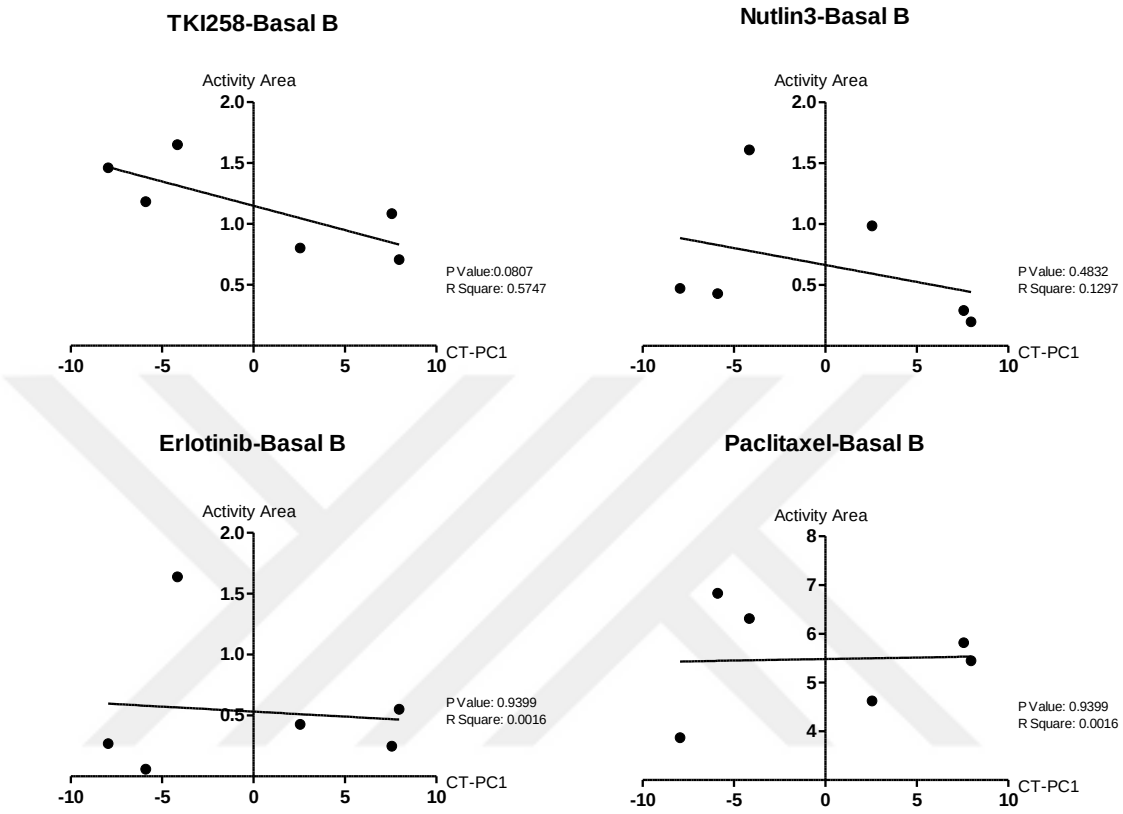


Figure 3.20: Other drug response correlation of CT-PC1 in Basal B cells. CT gene expression does correlate with drug response values of any other anticancer agents. Drug response values for HDQ-P, MDA-MB-157, Hs 578T, MDA-MB-436, BT-549 and HCC1395 Basal B cell lines were taken from CCLE database which contains pharmacologic profiles for 24 anticancer drugs across 504 cell lines. 28 of these cell lines are derived from breast, 7 of 28 are Basal B subtype-breast cancer cell lines. Pearson correlation analysis was performed by using drug response values (Activity area) and CT gene expression score calculated by first principal component analysis.

3.3.2 *In Vitro* Validation

To validate *in silico* findings, eight cell lines were treated with Panobinostat to measure its drug cytotoxicity (Figure 3.21). Then, their CT gene expression was measured by qRT-PCR to validate *in silico* findings in which CT gene expression can predict drug response (Figure 3.22).

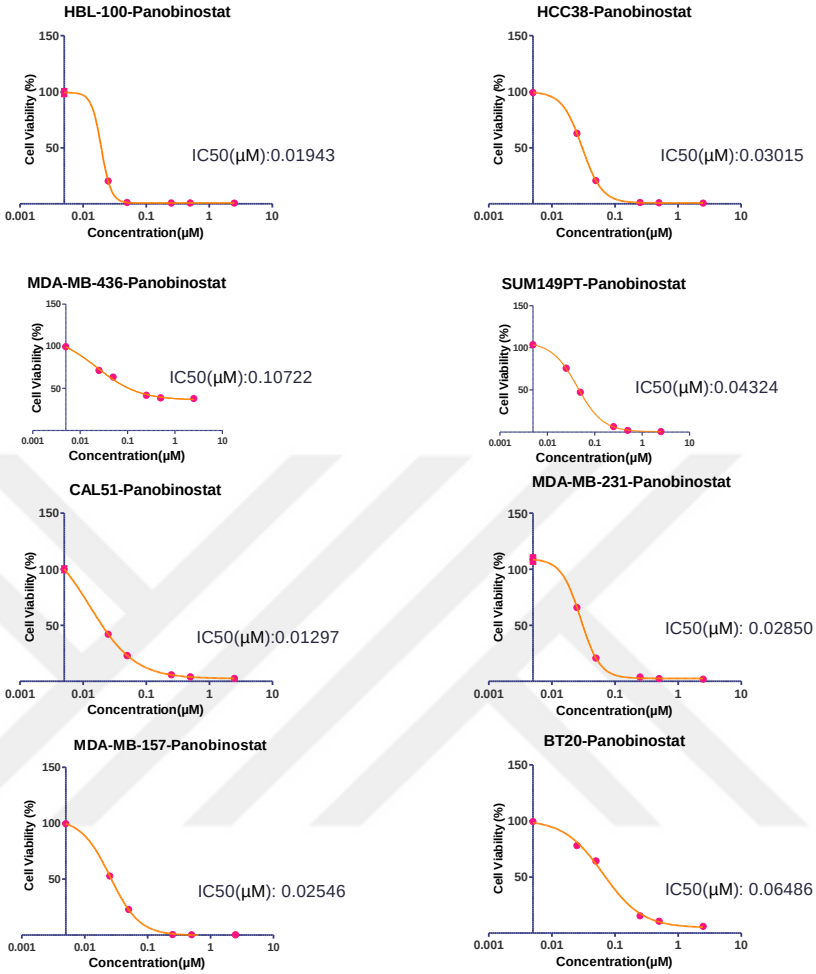


Figure 3.21: Percent cell viability curves for Panobinostat with Basal B cells. Basal B subtype breast cancer cell lines except MDA-MB-436 are highly sensitive to Panobinostat treatment. X-axis shows the concentrations of Panobinostat used in cell cytotoxicity experiments while y-axis shows percentage of cell viability in different concentration of drug. 5000 cells in 96-well plates were treated with 6 different concentrations as 2.5, 0.5, 0.25, 0.05, 0.025, 0.005 μM. Error bars represent median with 95% confidence interval.

Most sensitive cell line to Panobinostat treatment is CAL-51. BT-20 was most resistant cell line to Panobinostat treatment after MDA-MB-436.

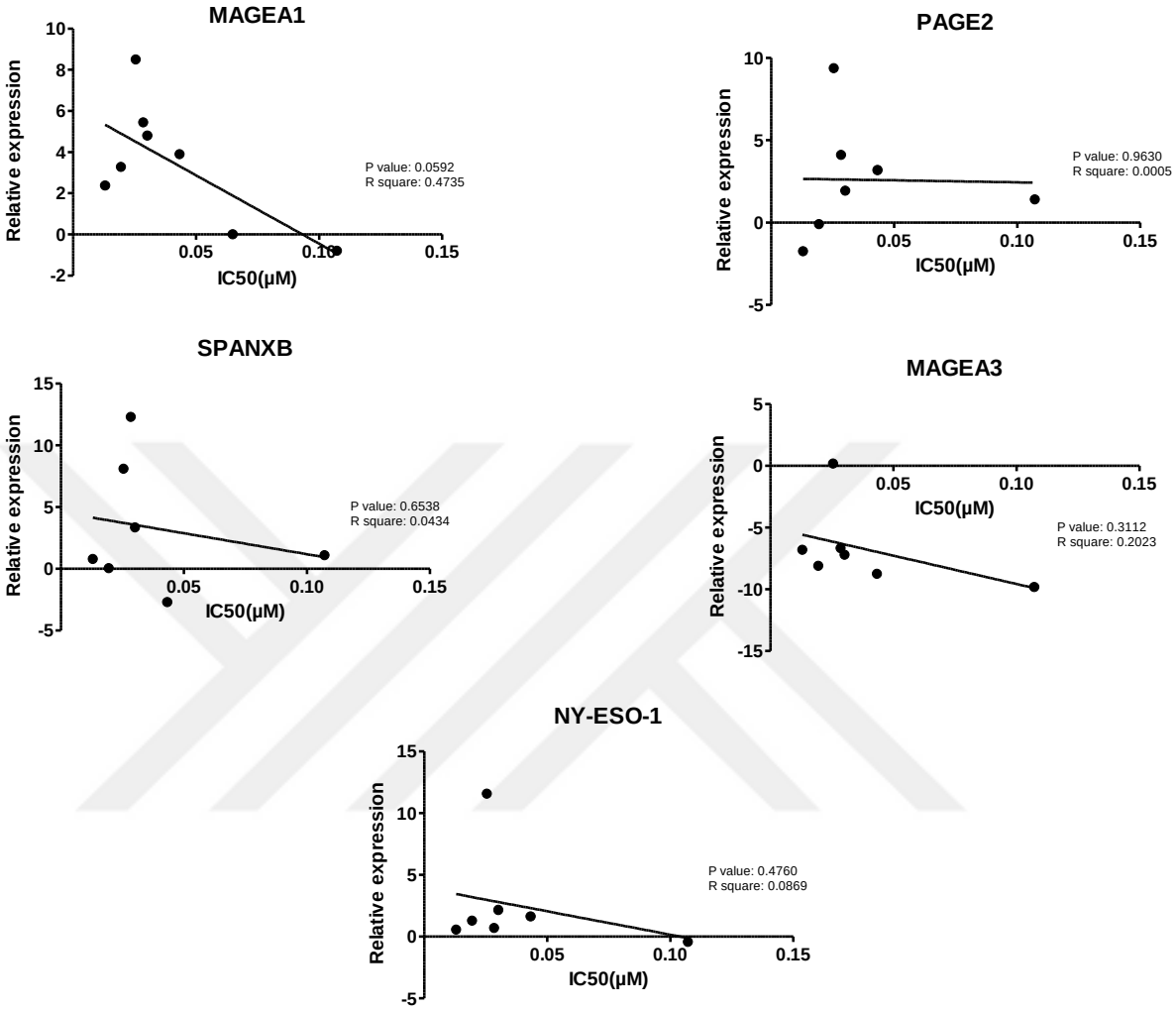


Figure 3.22: Drug response correlation of some CT gene expressions in Basal B cells. Of CT genes, MAGEA1 expression and Panobinostat cytotoxicity in Basal B breast cancer cell lines is borderline significant. X-axis shows the IC50 values for Panobinostat cytotoxicity while y-axis shows relative expression of represented genes in Basal B breast cancer cell lines. Relative gene expression is normalized according to expression value in BT20 cell line.

Similar pattern with *in silico* data was observed with *in vitro* validation experiments. We observed that CT-High cells are more sensitive to Panobinostat treatment. However, drug response and CT gene expression did not correlate well compared to *in silico* data.

4. DISCUSSION AND CONCLUSION

One of the hallmarks in tumorigenesis is aberrant expression pattern of genes which is resulted from global changes in epigenetic landscape [84]. Tissue-specific patterns of DNA methylation are profoundly dysregulated in cancer, aberrant hypermethylation is observed in tumor suppressor genes, and on the other hand, a variety of sequences including repetitive sequences is highly hypomethylated [85,86]. While silent in most healthy tissues, cancer testis antigen genes are frequently re-activated by promoter DNA hypomethylation and other epigenetic mechanisms in a wide range of cancer types [87]. This is an interesting finding, as CT genes contain more than 100 genes, some which are homologous because they are closely related genes (families), like MAGE-A, MAGE-B, NY-ESO, SSX, SPAN-X families, there is no sequence similarity between these genes that can easily explain why all undergo hypomethylation in cancer. While hypomethylation is common to tumors, it is not observed in all cells or tumors. As CT gene expression is directly related to tumorigenesis-associated hypomethylation, then this implies that a subtype of tumors, defined by CT gene upregulation, represent those in which a common mechanism, that is related to tumorigenesis is activated. So, studying the mechanism that causes this specific expression pattern of CT genes may reveal the epigenetic mechanisms which coordinate hypomethylation in promoter region of CT, and therefore elucidate a yet unknown mechanism that is altered in many cancer cells.

Both hypermethylation and hypomethylation can be detected within the same tumor cells [86]. This reflects indicates that these two modifications area region specific, and that such a region specific epigenetic aberration which resulted in de-repression of CT genes. To study region specific epigenetic mechanisms, we wanted to define the boundaries between a CT containing region, and one where opposite changes occurred simultaneously in cancer. we determined CT genes which are activated in cancer while their neighbor genes are repressed in cancer indicating separate epigenetic patterns. Once identified, we asked if these non-CT genes, that neighbor CT genes are whether they are tumor suppressor genes and whether their expression is regulated by DNA

methylation. Our group, however, previously showed that their ectopic expressions did not affect cell viability and that their methylation levels did not explain their expression pattern. Nevertheless, Therefore, we have tried to elucidate their control mechanism by studying their expression in on tumor and matched healthy tissues rather than healthy tissues and cancer cell lines, we asked if we could show the inverse association of their expression with CT genes, as had been shown earlier for cell lines. mRNA expression studies showed that only 3 of 8 tumor and matched healthy tissues have expected expression patterns. As the one reason for our inability to show an inverse expression pattern in these tissues can be the heterogeneous nature of tumor cells. Then, we decided to study their expression pattern repeat our analysis in homogeneous samples such as cancer cell lines. However, a clear, inverse relationship in expression pattern of CT and CT proximal genes could not be observed in cancer cell lines either. This indicates that the SPAN-X/ALAS2 and PAGE2/CDR1 models are probably not good models to study region specific epigenetic mechanism which control CT gene expression.

To find an answer to how CT genes are regulated, we decided to generate another model based on subgrouping cancer cell lines into CT-High, CT-Int and CT-Low types, reflecting the distribution of CT gene expression in tumor. We then compared gene expression differences among these groups in hope of obtaining clues that could explain the differential expression pattern of these genes.

We used three types of cancer cell lines, colon, breast and skin, reflecting CT-poor CT-moderate and CT-rich tumors, respectively. We hypothesized that categorizing cancer cell lines into CT-High, CT-Int and CT-low groups we may reveal the genes which coordinate epigenetic changes in the promoter regions of CT genes. Analyses were done by statistical tests to find differentially expressed genes between CT-High and CT-Low cell lines. Although there are differentially expressed genes within same cancer types, common genes among colon, breast and skin cancer were only CT genes. The reason why we could not identify genes that play active roles in inducing epigenetic changes by this approach may have been the subtle differences in the expression of these genes. Or, such differences might happen post-transcriptionally.

Alternatively, genes and mechanisms which coordinate epigenetic aberrations in the promoter of the CT genes can be different among different cancer types. Finally, it is known that epigenetic reprogramming occurs during epithelial-to-mesenchymal transition, mesenchymal or epithelial phenotype of the cell can cause de-repression of CT genes by different epigenetic mechanisms [36]. Furthermore, there are controversial findings about distribution of CT gene expression in epithelial or mesenchymal phenotype. Some studies claims that cells with invasive and mesenchymal phenotype have high levels of CT gene expression [33, 89]. On the other hand, many studies show that CT gene expression mainly associates with epithelial phenotype of cells [90,91]. Also, our recent findings suggest that SPANX-B and PAGE-2 genes are upregulated during mesenchymal-to epithelial transition with the increase in cytosine 5-hydroxymethylation levels in their CpG residues and dissociation of repressor proteins HP1 and EZH2 from their promoter-proximal regions [36]. So, these findings suggest studying epigenetic mechanism behind CT gene expression might differ based on whether the cell is transitioning from a mesenchymal to a more epithelial phenotype or the reverse.

We therefore categorized cancer cell lines into epithelial/CT-High, epithelial/CT-Low, mesenchymal/CT-high and mesenchymal/CT-Low for colon, breast and skin cancer. This categorization led us to discover the distribution of CT gene expression within cells with different phenotypes and change our previous hypothesis. Based on the findings in the study with dynamic mesenchymal-to-epithelial transition model, we proposed an EMT window in which CT genes are expressed between most epithelial and mesenchymal states, but not in these themselves [36]. However, our current data shows that tumor cells express CT genes when they have a more epithelial or mesenchymal phenotype. In other words, we think CT genes are expressed among epithelial cells that are more epithelial and among mesenchymal cells if they are more mesenchymal. This can explain the controversial findings on relation of CT gene expression in either epithelial or mesenchymal phenotype reported in the literature.

Thus, our finding verifies both expressions of CT genes associated with migratory phenotype and CT genes being part of the epithelial phenotype.

Our categorization of cancer cell lines based on their phenotype and CT expression status showed that breast and colon cancer cell lines have similar distribution in contrast to skin cancer cell lines. Skin cancer cell lines have a more mesenchymal phenotype with heterogeneous distribution of CT gene expression, containing both CT-High and CT-Low cells. Mesenchymal nature of melanoma was also shown before [92]. Within EMT defined subgroups, we tried to find differentially expressed genes between CT-High and CT-Low groups in epithelial or mesenchymal phenotype by statistical tests. When we compared significant genes identified in the separate tests for epithelial and mesenchymal phenotypes, we could not identify common genes other than CT genes. This suggested that genes or epigenetic control mechanisms which coordinate expression of CT genes in each phenotype are different. To identify such mechanisms, we performed gene set enrichment analysis with breast cancer cell lines. With gene set enrichment analysis, we could identify genes and epigenetic mechanisms in each phenotype separately by revealing enrichment of gene sets between CT-High and CT-Low groups.

Gene set enrichment analysis showed that there are different as well as common gene sets which are enriched in CT-High cells within both epithelial and mesenchymal phenotype. In mesenchymal cells, genes which are regulated by hippo pathway are enriched in CT-High/mesenchymal cells. Hippo pathway is thought to be required maintenance of tumor initiation capacities in breast CSCs [93]. This analysis reinforces our finding which is high levels of CT gene expression in more mesenchymal cells; even in CSCs. GSEA also suggested that essential epigenetic repressors BMI1 and PCGF2 proteins may have a role in cells with either mesenchymal or epithelial phenotypes. BMI1 is the component of polycomb repressive complex 1 and PCGF2 is similar to polycomb group repressors [94]. This reminded us our recent findings in which we showed dissociation of polycomb repressive complex 2 protein EZH2 from promoter proximal regions of CT genes while CT genes are upregulating. Genes

upregulated following transcriptional repressor proteins RB1 and RBL1 knockout were highly enriched in CT-High group of mesenchymal breast cancer cell lines. This may also suggest that RB1 and RBL1 have a role in regulation of CT gene expression. Genes related with MAPK-ERK and PI3K pathways were also enriched in CT-High epithelial as well as mesenchymal breast cancer cells indicating complexity in epigenetic control mechanisms of CT genes. On the other hand, genes downregulated following knockdown of MEK, EGFR and RAF1 were highly enriched in CT-Low epithelial breast cancer cell lines strongly suggesting that EGFR signaling pathway has a role in controlling CT gene expression among CT-high epithelial cells. Additionally, genes related to KRAS dependency and genes upregulated in cells overexpressing MYC gene are highly enriched in CT-Low mesenchymal breast cancer cells.

Previous findings and our results suggested that CT genes are highly expressed in triple-negative breast cancer and Basal B breast cancer cell lines. However, clinical relevance of CT genes in basal-like tumors remains largely unknown. To elucidate clinical significance of CT gene expression in basal-like tumors, we investigated any correlation between CT gene expression and drug response by analyzing CCLE and CGP drug databases. Although there are some good correlations, we chose Panobinostat, a pan-HDAC inhibitor. Histone deacetylase inhibitors (HDACi) are group of anti-cancer compounds [95, 96]. Sensitivity of cancer cells and resistance of healthy cells to HDACi show their ability to effect multiple epigenetic changes in cancer cells [97]. It has been shown that HDACi can modulate acetylation status not only histones but also variety of proteins, resulting in to be effective on growth, survival and differentiation [98, 99]. Effects of HDACi on cancer cell lines include cell morphology changes, activation of tumor suppressors, inactivation of oncogenes, induction of apoptosis, reduction in angiogenesis, and cell cycle arrest [100, 101]. Vorinostat, SAHA and Panobinostat are some of these HDACis. Panobinostat is termed as pan-HDAC indicating its activity against Class I, Class II and Class IV histone deacetylase enzymes [102]. There are some studies showing its potent inhibitory activity on some hematological malignancies [103-105]. Some recent studies also represented its activity

on solid tumors like small cell lung cancer, thyroid cancer with particular efficacy [106, 107]. Promising activity of HDACis on cancer cells accelerated clinical trials, using as single agent or in combination [96, 108-110]. Activity of Panobinostat on TNBC subtypes of breast cancers cells, MDA-MB-157, MDA-MB-231, MDA-MB468, and BT549 was also investigated. Of these, MDA-MB-468 was the most resistant. Basal B cell lines, MDA-MB-157, MDA-MB-231and BT549, were significantly responsive to nanomolar concentrations of Panobinostat. Tumor growth and progression of MDA-MB-157 and BT549 xenograft models, was also inhibited with low amounts of Panobinostat [111].

Concordant with *in silico* data, we observed that some CT-High Basal B cancer cell lines were more responsive to Panobinostat treatment. However, other CT-High Basal B cancer cell lines presented similar drug response with CT-low Basal B cancer cell lines. The reason for similar drug response correlation in some CT-High and CT-low triple-negative breast cancer cell lines can be inconsistency between measured CT mRNA expression and calculated CT-first principal component values. In principal component analysis, we have used expression values of 80 CT genes to generate first principal component values of cells. On the other hand, we have tried to correlate drug response with individual CT gene expression. Measurement of expression of more CT genes and calculation of PC1with these values may give results concordant correlations.

5. FUTURE PERSPECTIVES

In this study, we established a good model to study CT gene expression by subgrouping cancer cell lines first by their EMT profile then according to their CT expression levels. We have used data for colon, breast and skin cancer cell lines; these analyses in this study can be expanded with other types of cancer. Analysis of different cancer types would result in elucidation of common or various epigenetic mechanisms which coordinate CT gene expression.

Different approaches have been tried to define the subgroups in terms of CT gene expression. Distribution of CT gene expression in cell lines showed the relevance of epigenetic reprogramming in epithelial/mesenchymal phenotype on regulation of CT gene expression. *In vitro* dynamic EMT models from different sources of cancer can be used to reveal genes which coordinate epigenetic mechanisms leading to de-repression of CT genes in cancer.

Additionally, GSEA done by CT-Low and CT-High cancer cell lines resulted in candidate genes and pathways which may have role in re-activation of CT genes in cancer. Functional studies by targeting these genes and pathways will be next stages of this study.

BIBLIOGRAPHY

1. Simpson, A.J., et al., Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer, 2005. 5(8): p. 615-25.
2. Scanlan, M.J., A.J. Simpson, and L.J. Old, The cancer/testis genes: review, standardization, and commentary. Cancer Immun, 2004. 4: p. 1.
3. Costa, F.F., K. Le Blanc, and B. Brodin, Concise review: cancer/testis antigens, stem cells, and cancer. Stem Cells, 2007. 25(3): p. 707-11.
4. Egger, G., et al., Epigenetics in human disease and prospects for epigenetic therapy. Nature, 2004. 429(6990): p. 457-63.
5. Jones, P.A. and P.W. Laird, Cancer epigenetics comes of age. Nat Genet, 1999. 21(2): p. 163-7.
6. Jones, P.A. and S.B. Baylin, The fundamental role of epigenetic events in cancer. Nat Rev Genet, 2002. 3(6): p. 415-28.
7. Feinberg, A.P., R. Ohlsson, and S. Henikoff, The epigenetic progenitor origin of human cancer. Nat Rev Genet, 2006. 7(1): p. 21-33.
8. Gronbaek, K., C. Hother, and P.A. Jones, Epigenetic changes in cancer. APMIS, 2007. 115(10): p. 1039-59.
9. Cheng, Y.H., E.W. Wong, and C.Y. Cheng, Cancer/testis (CT) antigens, carcinogenesis and spermatogenesis. Spermatogenesis, 2011. 1(3): p. 209-220.
10. Almeida, L.G., et al., CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res, 2009. 37(Database issue): p. D816-9.
11. Coulie, P.G., et al., Genes coding for tumor antigens recognized by human cytolytic T lymphocytes. J Immunother Emphasis Tumor Immunol, 1993. 14(2): p. 104-9.
12. Wang, C., et al., Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. Nat Commun, 2016. 7: p. 10499.
13. Fratta, E., et al., The biology of cancer testis antigens: putative function, regulation and therapeutic potential. Mol Oncol, 2011. 5(2): p. 164-82.
14. Caballero, O.L. and Y.T. Chen, Cancer/testis (CT) antigens: potential targets for immunotherapy. Cancer Sci, 2009. 100(11): p. 2014-21.
15. Hofmann, O., et al., Genome-wide analysis of cancer/testis gene expression. Proc Natl Acad Sci U S A, 2008. 105(51): p. 20422-7.
16. Gure, A.O., et al., Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer. Clin Cancer Res, 2005. 11(22): p. 8055-62.
17. Scanlan, M.J., et al., Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. Immunol Rev, 2002. 188: p. 22-32.
18. De Smet, C., A. Loriot, and T. Boon, Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells. Mol Cell Biol, 2004. 24(11): p. 4781-90.

19. Sigalotti, L., et al., Intratumor heterogeneity of cancer/testis antigens expression in human cutaneous melanoma is methylation-regulated and functionally reverted by 5-aza-2'-deoxycytidine. Cancer Res, 2004. 64(24): p. 9167-71.
20. Pfeifer, G.P. and T.A. Rauch, DNA methylation patterns in lung carcinomas. Semin Cancer Biol, 2009. 19(3): p. 181-7.
21. De Smet, C. and A. Loriot, DNA hypomethylation in cancer: epigenetic scars of a neoplastic journey. Epigenetics, 2010. 5(3): p. 206-13.
22. Luetkens, T., et al., Expression, epigenetic regulation, and humoral immunogenicity of cancer-testis antigens in chronic myeloid leukemia. Leuk Res, 2010. 34(12): p. 1647-55.
23. James, S.R., et al., DNA methylation and nucleosome occupancy regulate the cancer germline antigen gene MAGEA11. Epigenetics, 2013. 8(8): p. 849-63.
24. Woloszynska-Read, A., et al., Intertumor and intratumor NY-ESO-1 expression heterogeneity is associated with promoter-specific and global DNA methylation status in ovarian cancer. Clin Cancer Res, 2008. 14(11): p. 3283-90.
25. Weiser, T.S., et al., Sequential 5-Aza-2 deoxycytidine-depsipeptide FR901228 treatment induces apoptosis preferentially in cancer cells and facilitates their recognition by cytolytic T lymphocytes specific for NY-ESO-1. J Immunother, 2001. 24(2): p. 151-61.
26. Wischnewski, F., K. Pantel, and H. Schwarzenbach, Promoter demethylation and histone acetylation mediate gene expression of MAGE-A1, -A2, -A3, and -A12 in human cancer cells. Mol Cancer Res, 2006. 4(5): p. 339-49.
27. Sun, F., et al., Combinatorial pharmacologic approaches target EZH2-mediated gene repression in breast cancer cells. Mol Cancer Ther, 2009. 8(12): p. 3191-202.
28. Cannuyer, J., et al., Epigenetic hierarchy within the MAGEA1 cancer-germline gene: promoter DNA methylation dictates local histone modifications. PLoS One, 2013. 8(3): p. e58743.
29. Rao, M., et al., Inhibition of histone lysine methylation enhances cancer-testis antigen expression in lung cancer cells: implications for adoptive immunotherapy of cancer. Cancer Res, 2011. 71(12): p. 4192-204.
30. Han, H., et al., Synergistic re-activation of epigenetically silenced genes by combinatorial inhibition of DNMTs and LSD1 in cancer cells. PLoS One, 2013. 8(9): p. e75136.
31. Hong, J.A., et al., Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. Cancer Res, 2005. 65(17): p. 7763-74.
32. Woloszynska-Read, A., et al., BORIS/CTCFL expression is insufficient for cancer-germline antigen gene expression and DNA hypomethylation in ovarian cell lines. Cancer Immun, 2010. 10: p. 6.
33. Cronwright, G., et al., Cancer/testis antigen expression in human mesenchymal stem cells: down-regulation of SSX impairs cell migration and matrix metalloproteinase 2 expression. Cancer Res, 2005. 65(6): p. 2207-15.
34. Chen, L., et al., Cancer/testis antigen SSX2 enhances invasiveness in MCF-7 cells by repressing ERalpha signaling. Int J Oncol, 2012. 40(6): p. 1986-94.

35. Chen, Y.T., et al., Cancer/testis antigen CT45: analysis of mRNA and protein expression in human cancer. Int J Cancer, 2009. 124(12): p. 2893-8.
36. Yilmaz-Ozcan, S., 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(9): p. e107905.
37. Novellino, L., C. Castelli, and G. Parmiani, A listing of human tumor antigens recognized by T cells: March 2004 update. Cancer Immunol Immunother, 2005. 54(3): p. 187-207.
38. Freitas, M., et al., Expression of cancer/testis antigens is correlated with improved survival in glioblastoma. Oncotarget, 2013. 4(4): p. 636-46.
39. Svobodova, S., et al., Cancer-testis antigen expression in primary cutaneous melanoma has independent prognostic value comparable to that of Breslow thickness, ulceration and mitotic rate. Eur J Cancer, 2011. 47(3): p. 460-9.
40. von Boehmer, L., et al., MAGE-C2/CT10 protein expression is an independent predictor of recurrence in prostate cancer. PLoS One, 2011. 6(7): p. e21366.
41. Zhou, X., et al., Heterogeneous expression of CT10, CT45 and GAGE7 antigens and their prognostic significance in human breast carcinoma. Jpn J Clin Oncol, 2013. 43(3): p. 243-50.
42. van Duin, M., et al., Cancer testis antigens in newly diagnosed and relapse multiple myeloma: prognostic markers and potential targets for immunotherapy. Haematologica, 2011. 96(11): p. 1662-9.
43. Dossus, L., et al., Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort. Int J Cancer, 2014. 134(8): p. 1871-88.
44. Connor, J., Alcohol consumption as a cause of cancer. Addiction, 2016.
45. Yang, T.O., et al., Birth weight and adult cancer incidence: large prospective study and meta-analysis. Ann Oncol, 2014. 25(9): p. 1836-43.
46. Inge, T.H., et al., The effect of obesity in adolescence on adult health status. Pediatrics, 2013. 132(6): p. 1098-104.
47. Biro, F.M. and M. Wien, Childhood obesity and adult morbidities. Am J Clin Nutr, 2010. 91(5): p. 1499S-1505S.
48. King, M.C. and A.G. Motulsky, Human genetics. Mapping human history. Science, 2002. 298(5602): p. 2342-3.
49. Singletary, S.E., Rating the risk factors for breast cancer. Ann Surg, 2003. 237(4): p. 474-82.
50. Welcsh, P.L., et al., BRCA1 transcriptionally regulates genes involved in breast tumorigenesis. Proc Natl Acad Sci U S A, 2002. 99(11): p. 7560-5.
51. Antoniou, A.C., et al., Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. Am J Hum Genet, 2008. 82(4): p. 937-48.
52. Evans, D.G., et al., Penetrance estimates for BRCA1 and BRCA2 based on genetic testing in a Clinical Cancer Genetics service setting: risks of breast/ovarian cancer quoted should reflect the cancer burden in the family. BMC Cancer, 2008. 8: p. 155.
53. Lord, S.J., et al., Breast cancer risk and hormone receptor status in older women by parity, age of first birth, and breastfeeding: a case-control study. Cancer Epidemiol Biomarkers Prev, 2008. 17(7): p. 1723-30.

54. Ma, H., et al., Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res, 2006. 8(4): p. R43.
55. Bladstrom, A., H. Anderson, and H. Olsson, Worse survival in breast cancer among women with recent childbirth: results from a Swedish population-based register study. Clin Breast Cancer, 2003. 4(4): p. 280-5.
56. MacMahon, B., et al., Age at first birth and breast cancer risk. Bull World Health Organ, 1970. 43(2): p. 209-21.
57. Lowe, C.R. and B. MacMahon, Breast cancer and reproduction. Lancet, 1970. 2(7683): p. 1137.
58. Usary, J., et al., Predicting drug responsiveness in human cancers using genetically engineered mice. Clin Cancer Res, 2013. 19(17): p. 4889-99.
59. Fillmore, C.M. and C. Kuperwasser, 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(2): p. R25.
60. Rudas, M., et al., Expression of MRP1, LRP and Pgp in breast carcinoma patients treated with preoperative chemotherapy. Breast Cancer Res Treat, 2003. 81(2): p. 149-57.
61. Langlands, F.E., et al., Breast cancer subtypes: response to radiotherapy and potential radiosensitisation. Br J Radiol, 2013. 86(1023): p. 20120601.
62. Haughian, J.M., et al., Maintenance of hormone responsiveness in luminal breast cancers by suppression of Notch. Proc Natl Acad Sci U S A, 2012. 109(8): p. 2742-7.
63. Eheman, C.R., et al., The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004. Cancer Epidemiol Biomarkers Prev, 2009. 18(6): p. 1763-9.
64. Genestie, C., et al., Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. Anticancer Res, 1998. 18(1B): p. 571-6.
65. Prat, A. and C.M. Perou, Deconstructing the molecular portraits of breast cancer. Mol Oncol, 2011. 5(1): p. 5-23.
66. Perou, C.M., et al., Molecular portraits of human breast tumours. Nature, 2000. 406(6797): p. 747-52.
67. Jonsson, G., et al., Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. Clin Cancer Res, 2010. 16(13): p. 3356-67.
68. Neve, R.M., et al., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell, 2006. 10(6): p. 515-27.
69. Tibshirani, R., et al., Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A, 2002. 99(10): p. 6567-72.
70. Sorlie, T., et al., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A, 2001. 98(19): p. 10869-74.

71. Chung, C.H., P.S. Bernard, and C.M. Perou, Molecular portraits and the family tree of cancer. Nat Genet, 2002. 32 Suppl: p. 533-40.
72. Perou, C.M., et al., Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A, 1999. 96(16): p. 9212-7.
73. Esmaeili, R., et al., AKAP3 correlates with triple negative status and disease free survival in breast cancer. BMC Cancer, 2015. 15: p. 681.
74. Lee, H.J., et al., Expression of NY-ESO-1 in Triple-Negative Breast Cancer Is Associated with Tumor-Infiltrating Lymphocytes and a Good Prognosis. Oncology, 2015. 89(6): p. 337-44.
75. Grigoriadis, A., et al., CT-X antigen expression in human breast cancer. Proc Natl Acad Sci U S A, 2009. 106(32): p. 13493-8.
76. Xu, X., et al., Overexpression of MAGE-A9 predicts unfavorable outcome in breast cancer. Exp Mol Pathol, 2014. 97(3): p. 579-84.
77. Abd-Elsalam, E.A. and N.A. Ismaeil, Melanoma-associated antigen genes: a new trend to predict the prognosis of breast cancer patients. Med Oncol, 2014. 31(11): p. 285.
78. Wong, P.P., et al., Identification of MAGEA antigens as causal players in the development of tamoxifen-resistant breast cancer. Oncogene, 2014. 33(37): p. 4579-88.
79. Irvin, W.J., Jr. and L.A. Carey, What is triple-negative breast cancer? Eur J Cancer, 2008. 44(18): p. 2799-805.
80. Lin, N.U., et al., Sites of distant recurrence and clinical outcomes in patients with metastatic triple-negative breast cancer: high incidence of central nervous system metastases. Cancer, 2008. 113(10): p. 2638-45.
81. Yao, J., et al., Tumor subtype-specific cancer-testis antigens as potential biomarkers and immunotherapeutic targets for cancers. Cancer Immunol Res, 2014. 2(4): p. 371-9.
82. Klijn, C., et al., A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol, 2015. 33(3): p. 306-12.
83. Nissan, A., et al., Colon cancer associated transcript-1: a novel RNA expressed in malignant and pre-malignant human tissues. Int J Cancer, 2012. 130(7): p. 1598-606.
84. Sharma, S., T.K. Kelly, and P.A. Jones, Epigenetics in cancer. Carcinogenesis, 2010. 31(1): p. 27-36.
85. Ehrlich, M., DNA methylation in cancer: too much, but also too little. Oncogene, 2002. 21(35): p. 5400-13.
86. Esteller, M., Epigenetics in cancer. N Engl J Med, 2008. 358(11): p. 1148-59.
87. Akers, S.N., K. Odunsi, and A.R. Karpf, Regulation of cancer germline antigen gene expression: implications for cancer immunotherapy. Future Oncol, 2010. 6(5): p. 717-32.
88. McDonald, O.G., et al., Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. Nat Struct Mol Biol, 2011. 18(8): p. 867-74.
89. Sigalotti, L., et al., Cancer testis antigens and melanoma stem cells: new promises for therapeutic intervention. Cancer Immunol Immunother, 2010. 59(3): p. 487-8.

90. Wallden, B., et al., Antimetastatic gene expression profiles mediated by retinoic acid receptor beta 2 in MDA-MB-435 breast cancer cells. BMC Cancer, 2005. 5: p. 140.
91. Gupta, P.B., et al., Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell, 2009. 138(4): p. 645-59.
92. Tan, T.Z., et al., Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO Mol Med, 2014. 6(10): p. 1279-93.
93. Cordenonsi, M., et al., The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell, 2011. 147(4): p. 759-72.
94. Wiederschain, D., et al., Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. Mol Cell Biol, 2007. 27(13): p. 4968-79.
95. Drummond, D.C., et al., Clinical development of histone deacetylase inhibitors as anticancer agents. Annu Rev Pharmacol Toxicol, 2005. 45: p. 495-528.
96. Liu, T., et al., Histone deacetylase inhibitors: multifunctional anticancer agents. Cancer Treat Rev, 2006. 32(3): p. 157-65.
97. Marks, P.A. and R. Breslow, Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. Nat Biotechnol, 2007. 25(1): p. 84-90.
98. Kikuchi, H., et al., Participation of histones, histone modifying enzymes and histone chaperones in vertebrate cell functions. Subcell Biochem, 2006. 40: p. 225-43.
99. Konstantinopoulos, P.A., M.V. Karamouzis, and A.G. Papavassiliou, Focus on acetylation: the role of histone deacetylase inhibitors in cancer therapy and beyond. Expert Opin Investig Drugs, 2007. 16(5): p. 569-71.
100. Vigushin, D.M. and R.C. Coombes, Histone deacetylase inhibitors in cancer treatment. Anticancer Drugs, 2002. 13(1): p. 1-13.
101. Lin, H.Y., et al., Targeting histone deacetylase in cancer therapy. Med Res Rev, 2006. 26(4): p. 397-413.
102. Xu, W.S., R.B. Parmigiani, and P.A. Marks, Histone deacetylase inhibitors: molecular mechanisms of action. Oncogene, 2007. 26(37): p. 5541-52.
103. Giles, F., et al., A phase I study of intravenous LBH589, a novel cinnamic hydroxamic acid analogue histone deacetylase inhibitor, in patients with refractory hematologic malignancies. Clin Cancer Res, 2006. 12(15): p. 4628-35.
104. Maiso, P., et al., The histone deacetylase inhibitor LBH589 is a potent antimyeloma agent that overcomes drug resistance. Cancer Res, 2006. 66(11): p. 5781-9.
105. Shao, W., et al., Activity of deacetylase inhibitor panobinostat (LBH589) in cutaneous T-cell lymphoma models: Defining molecular mechanisms of resistance. Int J Cancer, 2010. 127(9): p. 2199-208.
106. Crisanti, M.C., et al., The HDAC inhibitor panobinostat (LBH589) inhibits mesothelioma and lung cancer cells in vitro and in vivo with particular efficacy for small cell lung cancer. Mol Cancer Ther, 2009. 8(8): p. 2221-31.
107. Catalano, M.G., et al., Cytotoxic activity of the histone deacetylase inhibitor panobinostat (LBH589) in anaplastic thyroid cancer in vitro and in vivo. Int J Cancer, 2012. 130(3): p. 694-704.
108. Fukutomi, A., et al., A phase I study of oral panobinostat (LBH589) in Japanese patients with advanced solid tumors. Invest New Drugs, 2012. 30(3): p. 1096-106.

109. Ellis, L., et al., Histone deacetylase inhibitor panobinostat induces clinical responses with associated alterations in gene expression profiles in cutaneous T-cell lymphoma. Clin Cancer Res, 2008. 14(14): p. 4500-10.
110. Rathkopf, D., et al., A phase I study of oral panobinostat alone and in combination with docetaxel in patients with castration-resistant prostate cancer. Cancer Chemother Pharmacol, 2010. 66(1): p. 181-9.
111. Tate, C.R., et al., Targeting triple-negative breast cancer cells with the histone deacetylase inhibitor panobinostat. Breast Cancer Res, 2012. 14(3): p. R79.

A APPENDIX

Supplementary Table 1.1: Differentially expressed transcripts between CT-High and CT-Low skin cancer cell lines. Two hundred twenty one transcripts were identified as differentially expressed.

Melanoma							
Gene Name	**p-value**	**Adjusted p-val**	**Direction**	**Gene Name**	**p-value**	**Adjusted p-val**	**Direction**
TPTE	1.1E-12	2.86937E-08	UP	ARF6	4.67883E-05	0.017195672	DOWN
LOC100653084	2.68E-12	3.49293E-08	UP	GABRG2	4.8643E-05	0.017629026	UP
LOC100291796	1.6E-11	1.39293E-07	UP	PGRMC1	4.89106E-05	0.017483179	DOWN
LOC100288568	2.83E-11	1.84449E-07	UP	FLJ36000	5.17108E-05	0.018234353	UP
LOC100508797	7.26E-11	3.78875E-07	UP	LOC100509302	5.4235E-05	0.018869426	UP
DSCR4	1.68E-10	7.32528E-07	UP	RBM20	5.45647E-05	0.018734363	UP
LINC00221	1.75E-10	6.5173E-07	UP	SPCS1	6.32289E-05	0.021427202	UP
TAG	3.86E-10	1.2576E-06	UP	OXGR1	6.4542E-05	0.02159179	UP
BAGE2	4.42E-10	1.28143E-06	UP	LOC100134091	6.95484E-05	0.022972087	UP
BAGE4	5.16E-10	1.34629E-06	UP	OR8A1	7.04088E-05	0.022682075	UP
BAGE3	5.2E-10	1.23298E-06	UP	TNNI3	7.0554E-05	0.022451655	UP
C22orf34	2.8E-09	6.08343E-06	UP	AVPR2	7.52565E-05	0.023659571	UP
DSCR8	3.31E-09	6.64318E-06	UP	JAK1	7.62281E-05	0.023679703	DOWN
MAGEA12	4.68E-08	8.72358E-05	UP	SLC35D1	7.72773E-05	0.023723208	DOWN
PAGE2B	6.84E-08	0.00011892	UP	DUX4L9	7.93925E-05	0.024089175	UP
FLJ45974	8.58E-08	0.000139911	UP	CSPG5	8.17181E-05	0.024509791	UP
LOC400643	1.78E-07	0.000272947	UP	CTAG1A	9.08334E-05	0.026934181	UP
MAGEA10	3.86E-07	0.000559993	UP	CT45A4	9.28142E-05	0.027212287	UP
MGC39584	4.28E-07	0.000588105	UP	OR8G5	9.43522E-05	0.027355856	UP
CTAG2	5.52E-07	0.00071979	UP	CT45A5	9.45071E-05	0.027099653	UP
PAGE5	7.08E-07	0.000879971	UP	ANKRD45	0.000101901	0.02890224	UP
DHH	7.38E-07	0.000875157	UP	LOC93432	0.000104498	0.029320103	UP
CSAG1	9.12E-07	0.001034238	UP	KC6	0.0001075	0.029841465	UP
CTAG1B	1.09E-06	0.00118793	UP	USP19	0.000110551	0.030365429	UP
XAGE1D	1.27E-06	0.001323327	UP	MAGEA1	0.000110689	0.030086598	UP
XAGE1E	1.27E-06	0.001275822	UP	ZNF595	0.000111574	0.030014618	UP
MAGEA10-	1.29E-06	0.001250625	UP	TCL6	0.000113982	0.030349522	UP
MAGEA11	1.33E-06	0.001236623	UP	LOC285696	0.000117574	0.030989619	DOWN
KPNA6	1.52E-06	0.001365477	DOWN	LOC100009676	0.00011898	0.031046716	UP
MAGEA3	1.74E-06	0.001516317	UP	LOC100505840	0.000125738	0.032485311	UP
LOC100507559	2.53E-06	0.00212984	UP	SLCO1A2	0.000125842	0.032193376	UP
MYH8	2.63E-06	0.002146799	UP	FRG1B	0.000130057	0.032948641	UP

MAGEA2	3.06E-06	0.002421451	UP	TAL1	0.000132838	0.033329497	UP
TMEM57	5.28E-06	0.0040525	DOWN	LRRK1	0.000136572	0.033940003	DOWN
LOC100505874	5.96E-06	0.004446473	UP	MIPOL1	0.000148937	0.036663891	DOWN
LOC100128737	6.85E-06	0.004962003	UP	RGS12	0.000154376	0.037647641	DOWN
NAA11	7.63E-06	0.005378865	UP	KCNH5	0.000154533	0.037336986	UP
MAGEA5	8.58E-06	0.005893333	UP	CD81	0.00015811	0.037850587	DOWN
CSMD1	9.71E-06	0.006494076	UP	SELK	0.000158247	0.037538966	UP
MKRN9P	1.01E-05	0.006572829	UP	PTPN7	0.000158409	0.037238972	UP
MYH1	1.11E-05	0.007080279	UP	LOC100507599	0.000160974	0.037504137	UP
CT45A6	1.31E-05	0.008152975	UP	CLEC2L	0.000165375	0.038188387	UP
LOC100508631	1.4E-05	0.008502678	UP	RERG	0.000165667	0.037920329	UP
MAGEA2B	1.55E-05	0.009178189	UP	MYH4	0.000166234	0.03771918	UP
PAGE1	1.56E-05	0.009017882	UP	SMYD1	0.000166543	0.037463483	UP
FLJ46257	1.66E-05	0.009443341	UP	PCF11	0.000166812	0.037203286	DOWN
ADAMTS20	1.68E-05	0.009334042	UP	LOC386758	0.00016754	0.037049055	UP
MAGEA6	1.73E-05	0.009414458	UP	CT45A1	0.000167634	0.036758358	UP
LOC170425	1.77E-05	0.009439515	UP	DHRS7	0.000168392	0.036616881	DOWN
MYH13	1.93E-05	0.010066394	UP	ZDHHC19	0.000168651	0.036370094	UP
ATP1B2	2.01E-05	0.010297154	UP	KLK2	0.000170639	0.036497212	UP
LOC100506881	2.01E-05	0.010105759	UP	TTC25	0.00017962	0.038105817	UP
LOC100505490	2.05E-05	0.010092087	UP	HSD17B3	0.00018515	0.038962132	UP
SYNC	2.05E-05	0.009907247	DOWN	LOC100652816	0.000185325	0.038686888	DOWN
LOC649395	2.12E-05	0.01005229	UP	LOC442028	0.000197026	0.040803241	UP
CSAG3	2.28E-05	0.010626712	UP	VENTXP1	0.000200983	0.041294839	UP
NFE2	2.63E-05	0.012051089	UP	FAM46D	0.000208629	0.042530968	UP
LOC100289097	2.7E-05	0.012141625	UP	MARCKS	0.000211016	0.042684044	DOWN
LOC100506433	2.71E-05	0.011971866	UP	C4orf39	0.000211828	0.042518806	UP
CSAG2	3.01E-05	0.013098247	UP	SH3GLB1	0.000232992	0.046409841	DOWN
SLCO1B1	3.03E-05	0.012954108	UP	LOC100652887	0.000233139	0.046087282	UP
LOC100505948	3.25E-05	0.013675527	UP	LOC100652863	0.000233316	0.045775575	UP
RBBP4	3.28E-05	0.01357518	DOWN	MAGEB2	0.000233632	0.045495468	UP
CCDC71	4.08E-05	0.016619757	UP	OSBPL10	0.000235425	0.045505013	UP
MAGEC1	4.08E-05	0.016365371	UP	RTKN	0.000238817	0.045821312	UP
LOC100507370	4.09E-05	0.016160265	UP	ANGPTL1	0.00024342	0.04636347	UP
METTL7B	4.17E-05	0.016228286	UP	NLRP4	0.00024719	0.046740455	UP
NPRL2	4.55E-05	0.017464991	UP	LOC100653166	0.00026152	0.049094354	UP
LOC100147773	4.64E-05	0.017528621	UP	ZNF204P	0.000268761	0.050093214	UP
C3orf37	4.64E-05	0.017296623	UP	MYH2	0.000268844	0.049753332	UP

Supplementary Table 1.2: Differentially expressed transcripts between CT-High and CT-Low breast cancer cell lines. Twenty eight transcripts were identified as differentially expressed.

Breast			
Gene Name	p-value	Adjusted p-val	Direction
MAGEA2	1.15998E-14	3.02685E-10	UP
MAGEA3	3.07576E-12	4.01294E-08	UP
CSAG1	1.07575E-11	9.35691E-08	UP
CSAG2	3.98315E-11	2.59841E-07	UP
MAGEA12	2.96276E-10	1.54621E-06	UP
MAGEA2B	4.88109E-10	2.12279E-06	UP
DGKB	6.64706E-08	0.000247783	UP
MAGEA6	1.86976E-07	0.00060987	UP
CSAG3	2.97013E-07	0.00086114	UP
GABRA3	5.82947E-07	0.001521142	UP
KCNMB2	2.75795E-06	0.006542351	UP
LOC100509302	5.98472E-06	0.013013778	UP
PAGE2B	9.91497E-06	0.019901629	UP
FLJ45974	1.02179E-05	0.019044697	UP
DDO	1.21381E-05	0.021115389	UP
PAGE1	2.76341E-05	0.045067705	UP
LINC00221	2.90714E-05	0.04462288	UP
LOC100288568	3.04533E-05	0.044147114	UP
MAGEB6	4.38796E-05	0.060262825	UP
ODZ1	6.66616E-05	0.086973404	UP
LOC100653084	7.71214E-05	0.095828801	UP
SSX1	8.02645E-05	0.095200951	UP
XAGE2	8.59707E-05	0.097535613	UP
PPP1R1C	9.13982E-05	0.09937269	UP
PAGE2	9.30829E-05	0.097156253	UP
SHANK3	9.49379E-05	0.095281106	DOWN
TAG	9.61549E-05	0.092928407	UP
CTAG1B	0.000115896	0.108006786	UP

Supplementary Table 1.3: Differentially expressed transcripts between CT-High and CT-Low colon cancer cell lines. Thirty three transcripts were identified as differentially expressed.

Colon			
Gene name	**p-value**	**Adjusted p-val**	**Direction**
MAGEA6	5.51935E-12	1.44022E-07	UP
MAGEA2	1.05318E-10	1.37408E-06	UP
CSAG2	5.33238E-09	4.6381E-05	UP

SSX1	8.92113E-09	5.8197E-05	UP
MAGEA3	1.61449E-08	8.42571E-05	UP
CSAG1	2.75711E-08	0.000119907	UP
GABRA3	8.27943E-08	0.000308633	UP
MAGEB6	1.38117E-07	0.000450504	UP
MAGEA12	3.2752E-07	0.000949591	UP
MKRN3	9.61992E-07	0.002510222	UP
ZNF606	1.68848E-06	0.004005378	UP
PAGE2B	2.81932E-06	0.006130607	UP
FLJ42875	1.29424E-05	0.025978366	DOWN
LOC441666	1.99779E-05	0.037236	UP
LOC440157	2.2824E-05	0.039704564	UP
ODZ1	3.90958E-05	0.063760409	UP
PAPPA2	4.26659E-05	0.065489605	UP
SMCR5	4.38423E-05	0.063556701	UP
ZNF350	4.52627E-05	0.062162328	UP
AACSP1	4.74251E-05	0.061875471	UP
F13A1	7.05035E-05	0.087605599	UP
TPTE2P6	7.1956E-05	0.08534639	UP
CDC37L1	7.36688E-05	0.083578874	DOWN
RANBP3L	7.65295E-05	0.083206647	UP
MAGEA2B	7.72548E-05	0.080635439	UP
TMEM9B	8.22085E-05	0.082505746	DOWN
TAS2R43	8.5326E-05	0.082462835	UP
LOC100291796	8.85104E-05	0.082485379	UP
ZNF550	0.00010679	0.096089252	UP
VN1R1	0.000109139	0.094928886	UP
BGN	0.000115923	0.097576895	UP
TOPORS	0.00011637	0.094892173	DOWN
MAGEA1	0.000124622	0.098542403	UP

Supplementary Table 1.4: Differentially expressed non-coding RNAs between CT-High and CT-Low skin cancer cell lines. Seventeen non-coding genes were identified as differentially expressed between CT-High and CT-Low skin cancer cell lines.

Melanoma			
geneID	**p-value**	**Adjusted p-val**	**Direction**
ENSG00000248783	2.13568E-10	2.60916E-06	UP
ENSG00000251363	3.49726E-09	4.27261E-05	UP
ENSG00000248103	5.04236E-09	6.16025E-05	UP
ENSG00000229131	9.77918E-09	0.000119472	UP
ENSG00000250453	1.53702E-08	0.000187778	UP
ENSG00000253642	1.5513E-07	0.001895218	UP

ENSG00000242781	2.93153E-07	0.003581447	UP
ENSG00000189229	3.22109E-07	0.003935201	UP
ENSG00000233515	1.16179E-06	0.014193585	UP
ENSG00000254302	1.30787E-06	0.015978204	UP
ENSG00000258754	2.07927E-06	0.025402444	UP
ENSG00000258038	2.27235E-06	0.027761314	UP
ENSG00000232765	2.44886E-06	0.029917708	UP
ENSG00000258028	2.75167E-06	0.033617131	UP
ENSG00000242828	2.89257E-06	0.035338531	UP
ENSG00000258688	3.27406E-06	0.039999179	UP
ENSG00000258476	3.53848E-06	0.043229597	UP

Supplementary Table 1.5: Differentially expressed non-codin RNAs transcripts between CT-High and CT-Low breast cancer cell lines. Nineteen non-coding genes were identified as differentially expressed between CT-High and CT-Low breast cancer cell lines.

Breast			
geneID	**p-value**	**Adjusted p-val**	**Direction**
ENSG00000224037	4.46613E-07	0.005456269	UP
ENSG00000230880	9.59556E-07	0.011722901	UP
ENSG00000203849	1.31478E-06	0.004382589	UP
ENSG00000232694	1.81146E-06	0.004528641	UP
ENSG00000255319	1.989E-06	0.003978006	UP
ENSG00000248138	2.64731E-06	0.004412176	UP
ENSG00000238261	3.98812E-06	0.005697311	UP
ENSG00000251003	5.14099E-06	0.006426241	UP
ENSG00000233080	5.3421E-06	0.005935665	UP
ENSG00000232274	7.01091E-06	0.00701091	UP
ENSG00000247735	7.75494E-06	0.007049944	UP
ENSG00000227674	8.45478E-06	0.007045649	UP
ENSG00000251026	1.01361E-05	0.007796989	UP
ENSG00000258556	1.51744E-05	0.010838877	UP
ENSG00000230850	1.73562E-05	0.011570809	UP
ENSG00000185044	2.59361E-05	0.016210068	UP
ENSG00000212569	5.121E-05	0.030123516	UP
ENSG00000257869	7.8042E-05	0.043356664	UP
ENSG00000249345	8.56253E-05	0.045065933	UP

Supplementary Table 1.6: Differentially expressed non-coding transcripts between CT-High and CT-Low colon cancer cell lines. Three non-coding genes were identified as differentially expressed between CT-High and CT-Low colon cancer cell lines.

Colon			
geneID	**p-value**	**Adjusted p-val**	**Direction**
ENSG00000224271	6.31722E-09	7.71775E-05	UP
ENSG00000230105	1.70499E-08	0.000208298	UP
ENSG00000225278	1.93176E-05	0.064392135	UP

Supplementary Table 1.7: Differentially expressed transcripts between CT-High and CT-Low breast epithelial cancer cell lines. Top 100 genes were listed below.

BREAST/EPITHELIAL							
Gene Name	**p-value**	**Adjusted p-val**	**Direction**	**Gene Name**	**p-value**	**Adjusted p-val**	**Direction**
CSAG2	4.74926E-15	1.23927E-10	UP	ZNF24	0.00052	0.265912	DOWN
MAGEA3	7.29397E-10	9.51644E-06	UP	C19orf38	0.000528	0.264886	DOWN
MAGEA2	1.35022E-09	1.17442E-05	UP	KCNMB2	0.000541	0.266203	UP
CSAG1	1.99443E-08	0.000130107	UP	SEL1L2	0.000541	0.261489	DOWN
MAGEA12	3.81686E-07	0.001991942	UP	IKBKB	0.000559	0.265406	DOWN
MAGEA2B	4.92254E-07	0.002140814	UP	PPM1A	0.000567	0.264108	DOWN
OMD	2.751E-06	0.010254948	DOWN	ERI2	0.000578	0.264457	DOWN
KAT6A	3.2275E-06	0.010527312	DOWN	N4BP2	0.000591	0.266017	DOWN
SDK1	3.42068E-06	0.00991769	DOWN	AGPAT6	0.000599	0.264769	DOWN
LOC440900	7.0148E-06	0.018304407	UP	SNTN	0.00061	0.26527	DOWN
DGKB	1.09511E-05	0.025978035	UP	LOC100292909	0.000624	0.266803	DOWN
CSAG3	2.73999E-05	0.059581153	UP	HUS1B	0.000625	0.26308	UP
MAGEA6	2.77462E-05	0.055693117	UP	BPTF	0.000659	0.272943	DOWN
TLR10	3.07536E-05	0.057320238	UP	HGSNAT	0.000676	0.275526	DOWN
TM7SF4	3.44238E-05	0.059883671	DOWN	CD38	0.000696	0.279372	UP
SLC39A6	4.5078E-05	0.073516636	DOWN	HES5	0.000745	0.294635	UP
LOC100507003	5.10903E-05	0.078420568	UP	ANKRD26P1	0.000764	0.29745	UP
GABRA3	6.76617E-05	0.098086964	UP	UBAP2L	0.000818	0.313843	DOWN
FCGR3A	7.51207E-05	0.103168344	DOWN	C21orf7	0.000857	0.32405	UP
IL5RA	7.82558E-05	0.1021004	DOWN	ANGEL2	0.000864	0.321946	DOWN
IARS2	9.68115E-05	0.120295214	DOWN	SUSD4	0.000865	0.318011	DOWN
C12orf66	0.000105709	0.125380044	DOWN	LOC653125	0.000875	0.317038	DOWN
HOOK3	0.000111125	0.126074261	DOWN	BACE2	0.00088	0.314445	UP
GOLGA8J	0.000111851	0.121610242	DOWN	MIR3173	0.000915	0.322629	DOWN
TXNDC16	0.00011202	0.116921801	DOWN	LOC339290	0.000949	0.330304	UP

LAMB1	0.000122349	0.12279123	UP	RPL23AP32	0.000996	0.341859	DOWN
LOC91149	0.000123142	0.119010186	DOWN	MEP1A	0.001018	0.344997	UP
FUT5	0.00012373	0.115307806	UP	FGFR1	0.001031	0.34506	DOWN
POLI	0.000128895	0.115978665	DOWN	KIAA0825	0.001144	0.377832	DOWN
FDPSL2A	0.000145598	0.126640775	UP	LOC389834	0.001149	0.374639	UP
DTNA	0.000208282	0.175319972	DOWN	RSPH3	0.001155	0.372081	DOWN
LOC100506030	0.00020883	0.170287936	DOWN	HCN2	0.001194	0.380058	DOWN
RAD50	0.000235881	0.186517782	DOWN	LOC100129744	0.001236	0.388619	DOWN
ADAMTS20	0.000247725	0.190121288	UP	CDRT1	0.001237	0.384237	UP
FAM135B	0.000259002	0.193097152	DOWN	WDR7	0.001247	0.382769	DOWN
SNX25	0.000259067	0.187780162	DOWN	MGC39584	0.001265	0.383846	UP
LOC100652922	0.000273212	0.192680567	DOWN	LOC285501	0.001283	0.384832	UP
LOC100653205	0.000273212	0.187610026	DOWN	LOC100652904	0.001287	0.381751	DOWN
HIST1H1A	0.000282735	0.189171193	UP	LOC100653142	0.001287	0.377462	DOWN
ADAMTS19	0.000287093	0.187284929	DOWN	SOAT2	0.001306	0.37867	DOWN
SMYD1	0.000291799	0.185712041	DOWN	DDO	0.001309	0.375364	UP
MKRN3	0.000358439	0.222693311	UP	BCL2L12	0.001325	0.37568	UP
ZNF337	0.000386932	0.234804767	DOWN	PFKFB4	0.00134	0.375873	DOWN
GALM	0.000389792	0.231164455	UP	PCNX	0.001355	0.37614	DOWN
WHSC1L1	0.000408591	0.236928133	DOWN	AKT1S1	0.001405	0.385952	UP
LOC100506397	0.000431254	0.244633367	DOWN	RIPPLY1	0.001471	0.399839	UP
MIR4692	0.000457384	0.253935685	DOWN	LOC340544	0.001521	0.409144	DOWN
ARHGEF25	0.000480758	0.261352181	DOWN	CETP	0.001534	0.408332	UP
RPL23AP7	0.000481802	0.256574185	UP	MTRNR2L6	0.001549	0.408285	UP
GPR155	0.000515717	0.269142367	DOWN	OR52E6	0.0016	0.417629	UP

Supplementary Table 1.8: Differentially expressed transcripts between CT-High and CT-Low breast mesenchymal cancer cell lines. Top 100 genes were listed below.

BREAST/MESENCHYMAL							
Gene Name	**p-value**	**Adjusted p-val**	**Direction**	**Gene Name**	**p-value**	**Adjusted p-val**	**Direction**
LOC100291796	4.20378E-08	0.001096933	UP	DOC2A	0.000367	0.187822	DOWN
LOC100129316	1.65853E-07	0.002163879	DOWN	PURB	0.000377	0.189092	UP
LOC100509445	2.89846E-07	0.002521081	DOWN	YBEY	0.000391	0.192282	DOWN
FLJ45974	5.09342E-07	0.003322694	UP	SLC13A5	0.000446	0.2155	DOWN
BAGE4	2.38906E-06	0.012468036	UP	UBQLNL	0.000474	0.224772	UP
BAGE3	2.38906E-06	0.01039003	UP	LOC440157	0.000477	0.222344	UP
MAGEA2	3.06988E-06	0.011443649	UP	SH3BP4	0.00049	0.224147	UP
LOC100505565	4.67688E-06	0.015254815	DOWN	DMRT2	0.0005	0.224841	DOWN

CHDH	1.43006E-05	0.041462217	DOWN	MAGEA3	0.000509	0.225022	UP
LCP2	1.78025E-05	0.046453767	UP	FAM114A2	0.000509	0.221573	UP
LRRC14B	1.90912E-05	0.045287816	DOWN	IFLTD1	0.000516	0.22053	UP
KRTDAP	2.24414E-05	0.048798819	DOWN	TCL6	0.000523	0.219999	UP
ANO3	3.00632E-05	0.060343832	UP	C6orf52	0.000545	0.225823	DOWN
MAGEA2B	3.19897E-05	0.059624299	UP	PRTFDC1	0.000571	0.232607	DOWN
TAG	5.02184E-05	0.087359868	UP	ACTB	0.000579	0.232539	UP
LOC100505767	5.17005E-05	0.0843171	DOWN	LOC284950	0.000582	0.229919	UP
LOC100132781	5.27628E-05	0.08098779	DOWN	AQP4	0.000611	0.238034	DOWN
GIF	5.96927E-05	0.086534503	DOWN	SLC25A2	0.000623	0.239074	DOWN
CEBPE	6.00464E-05	0.082465836	DOWN	LOC255411	0.000639	0.241652	DOWN
LOC643696	6.0521E-05	0.078961719	DOWN	LOC100506646	0.000648	0.241462	DOWN
GOLGA6L10	6.0521E-05	0.075201637	DOWN	LOC100653054	0.000648	0.238061	DOWN
GOLGA6B	6.20893E-05	0.073643546	DOWN	CEACAM8	0.000651	0.236043	DOWN
TMEM236	6.26172E-05	0.07104058	DOWN	LOC100505588	0.000659	0.235552	DOWN
FETUB	6.9741E-05	0.075825868	DOWN	LOC147646	0.000677	0.238765	DOWN
BAGE2	9.80532E-05	0.10234399	UP	LOC100506890	0.000683	0.237539	DOWN
CD151	0.000100552	0.100915657	UP	FAM122B	0.000692	0.237652	DOWN
ERMN	0.000111977	0.10821983	UP	SPDYE8P	0.000703	0.238142	DOWN
LOC100505550	0.000121329	0.113070425	DOWN	FAM86C2P	0.000704	0.23554	DOWN
SPINK2	0.00012328	0.11092639	DOWN	PHKA1-AS1	0.000711	0.234968	DOWN
PRR18	0.000125002	0.108726656	DOWN	OR6B2	0.000712	0.232099	DOWN
DCAF8L2	0.000172958	0.145586278	UP	DAPL1	0.000716	0.230648	DOWN
LOC285556	0.000198332	0.161727059	UP	EGFLAM	0.00072	0.228996	UP
ZNF33A	0.000208197	0.164627114	DOWN	PINK1	0.000723	0.227189	UP
LOC100287195	0.000212646	0.163199742	DOWN	PDCL2	0.000728	0.226052	DOWN
TFDP3	0.000227729	0.169781908	DOWN	CFHR3	0.000729	0.223674	UP
RIC8A	0.000237712	0.172301818	UP	MIR4645	0.000737	0.223623	DOWN
CARD17	0.000242286	0.170870274	UP	MIR645	0.000747	0.224073	DOWN
LOC441239	0.00025779	0.177020357	DOWN	TRNF	0.000747	0.221596	DOWN
RIPPLY2	0.000260922	0.174576886	DOWN	SLC16A14	0.000751	0.220308	DOWN
MFAP3	0.000264467	0.172525308	UP	LOC440292	0.000753	0.218436	DOWN
OBSCN	0.000268264	0.170733995	DOWN	MUC15	0.000786	0.225258	UP
LOC643733	0.000295296	0.183463306	UP	FLJ46257	0.000787	0.223099	UP
LOC646324	0.000303558	0.184210442	UP	DBNL	0.000789	0.221324	UP
MIR554	0.000306484	0.181758754	DOWN	ANKRD30BP2	0.000808	0.224376	DOWN
GRM3	0.000315318	0.182842315	UP	SOSTDC1	0.000814	0.22357	DOWN
HRK	0.000325258	0.18450605	DOWN	LOC79999	0.000847	0.230213	DOWN
PADI6	0.00033884	0.188121158	DOWN	TIPIN	0.000854	0.229639	DOWN
LOC100652797	0.000353643	0.192249105	DOWN	LOC100507468	0.000855	0.227721	DOWN

LOC100653256	0.000353643	0.188325654	DOWN	HTR3C	0.000856	0.225511	UP
PAGE1	0.000356073	0.185827167	UP	LOC100509100	0.000857	0.223578	DOWN

Supplementary Table 1.9: Differentially expressed transcripts between CT-High and CT-Low colon epithelial cancer cell lines. Top 100 genes were listed below.

COLON/EPITHELIAL							
Gene name	**p-value**	**Adjusted p-val**	**Direction**	**Gene name**	**p-value**	**Adjusted p-val**	**Direction**
MAGEA6	8.46398E-11	2.20859E-06	UP	AMACR	7.54E-05	0.039327856	DOWN
MAGEA2	5.29413E-08	0.000690725	UP	WRNIP1	7.74E-05	0.03958117	UP
MEIS1	6.63186E-08	0.000576839	UP	FKSG29	8.12E-05	0.040755507	UP
LOC100506303	8.03296E-08	0.00052403	UP	EPGN	8.71E-05	0.042898772	UP
LRRC25	1.47938E-07	0.000772061	UP	PAGE2B	8.76E-05	0.042331448	UP
LOC440905	2.11259E-07	0.000918765	UP	ZNF577	8.93E-05	0.042351623	UP
MIR203	3.04509E-07	0.001135124	UP	ZNF841	9.82E-05	0.045779831	UP
LOC440157	3.26674E-07	0.00106553	UP	L3MBTL3	1E-04	0.045777039	DOWN
CSAG2	4.13336E-07	0.001198399	UP	LOC100132831	0.000104	0.046631632	UP
LOC100653149	4.43288E-07	0.001156716	UP	SNORD7	0.000104	0.046182769	UP
IL9R	4.56159E-07	0.001082093	UP	LOC642366	0.000111	0.048239303	UP
ILDR2	7.42188E-07	0.001613887	DOWN	LOC100507760	0.000113	0.048460383	DOWN
ATP2B3	8.94773E-07	0.001796016	UP	GPR20	0.000116	0.048734896	UP
MKRN3	1.0544E-06	0.001965246	UP	BARX1	0.000116	0.048241662	UP
CABP2	1.09668E-06	0.001907784	UP	ATHL1	0.000122	0.049935803	DOWN
TAS2R43	1.51936E-06	0.002477883	UP	PCDHB4	0.000134	0.053693936	UP
MAGEA3	2.95331E-06	0.004533152	UP	LOC728728	0.00014	0.0544734	UP
PLAG1	4.20128E-06	0.006090452	UP	RNU5E-1	0.000181	0.069371799	UP
CECR7	5.41267E-06	0.007433588	UP	AACSP1	0.000187	0.07058661	UP
LOC728648	5.94365E-06	0.007754678	UP	LOC100288814	0.000192	0.071629389	DOWN
FLJ39632	5.9979E-06	0.007452816	UP	ADAMTS2	0.000198	0.07259606	UP
MIR146A	6.35882E-06	0.007542138	DOWN	SNORD3B-2	0.0002	0.072485069	DOWN
TMEM9B	6.42116E-06	0.007284947	DOWN	PCDHB3	0.000204	0.073065959	UP
RRM1	6.85259E-06	0.007450483	DOWN	LOC283089	0.000218	0.076888847	UP
CSAG1	7.1794E-06	0.007493574	UP	FOXS1	0.000252	0.087749061	UP
FLJ42393	1.02269E-05	0.01026388	UP	LOC100507387	0.000254	0.087188942	DOWN
ZNF492	1.09519E-05	0.010584415	UP	ANGPTL7	0.000256	0.086737522	UP
MAGEA12	1.41698E-05	0.013205236	UP	HEY2	0.000264	0.08845973	UP
ZNF350	1.68303E-05	0.015143762	UP	C11orf34	0.000269	0.088810683	UP
LOC100294020	1.94783E-05	0.016942266	UP	LOC100507359	0.00027	0.088053808	UP

LOC100293748	2.25206E-05	0.018956561	UP	C9orf153	0.000271	0.087252948	UP
DPPA5	2.32864E-05	0.018988574	UP	SMCR5	0.000275	0.087421954	UP
ZNF550	3.50951E-05	0.027750656	UP	TMEM98	0.000277	0.087178112	DOWN
LOC100507487	3.8063E-05	0.029212241	UP	MIR941-1	0.000298	0.092424031	UP
MIR3545	3.82434E-05	0.028512057	UP	NLRP3	0.000305	0.093741688	UP
SSX1	3.84833E-05	0.027894	UP	MIR4467	0.000313	0.095081552	UP
FAM65A	4.00665E-05	0.028256642	UP	USP32P2	0.000314	0.094243734	UP
TAS2R30	4.2946E-05	0.029490312	UP	SCAND3	0.000319	0.094448168	UP
TEX13B	4.45451E-05	0.029804065	UP	MIR302A	0.00033	0.09689429	UP
CD99	5.00165E-05	0.032628264	DOWN	C2orf53	0.000338	0.097984316	UP
C1QTNF3-AMACR	5.01282E-05	0.031903566	DOWN	HIST1H3H	0.00036	0.103154929	UP
PRSS38	5.22196E-05	0.032443295	UP	CEACAM22P	0.000363	0.103067603	UP
FAM205A	5.25771E-05	0.031905761	UP	RAB40AL	0.000368	0.103299665	UP
GAGE2A	5.31892E-05	0.031543598	UP	LOC100130698	0.000371	0.102911602	UP
SIRPA	5.36631E-05	0.031117433	UP	ZNF263	0.00039	0.107245946	UP
MAGEB6	5.55085E-05	0.031487806	UP	ATP1A3	0.000405	0.110084528	DOWN
PCDHB15	5.71396E-05	0.031723447	UP	RAB7L1	0.000407	0.109432315	DOWN
MGC23284	5.74388E-05	0.031225162	UP	SOX1	0.000408	0.108688179	DOWN
OSM	5.96066E-05	0.031742353	UP	JARID2-AS1	0.000414	0.108990549	UP

Supplementary Table 1.10: Differentially expressed transcripts between CT-High and CT-Low colon mesenchymal cancer cell lines. Top 100 genes were listed below.

COLON/MESENCHYMAL							
Gene name	**p-value**	**Adjusted p-val**	**Direction**	**Gene name**	**p-value**	**Adjusted p-val**	**Direction**
GSTT1	1.55453E-06	0.040563866	DOWN	ST8SIA4	0.000939532	0.480708809	UP
SGPL1	9.58458E-06	0.125050043	DOWN	HIF1AN	0.000941715	0.472560044	DOWN
KRTAP2-4	1.05753E-05	0.091983992	DOWN	LOC731424	0.000942953	0.464253223	UP
ADARB2	3.40223E-05	0.221944768	UP	SLC2A6	0.000973974	0.470645741	DOWN
LINC00244	6.62791E-05	0.345897385	UP	ALS2CL	0.000975157	0.462649965	DOWN
RBM11	9.20257E-05	0.400219859	UP	GPC5	0.000983215	0.458142916	UP
ARHGEF26	9.25322E-05	0.344933667	UP	PI15	0.001018996	0.46648568	UP
LOC286367	0.000102284	0.333625121	DOWN	UBP1	0.001024809	0.461058136	DOWN
MAGEA2	0.000140184	0.406441497	UP	SLC16A3	0.00103199	0.456419338	DOWN
SCARNA3	0.000165275	0.431267392	UP	ISG15	0.001032022	0.448826415	DOWN
SLC5A10	0.00016608	0.393972838	DOWN	MIR4737	0.001044748	0.44691251	UP
PDIA6	0.00017175	0.373469887	UP	AMOT	0.001076409	0.453029468	UP
ZNF2	0.00018258	0.366479662	UP	CCDC121	0.001096984	0.454360228	UP
PHGDH	0.000199055	0.371009639	UP	ARHGEF26-AS1	0.0011015	0.449102177	UP

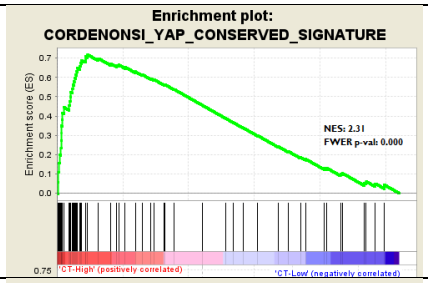
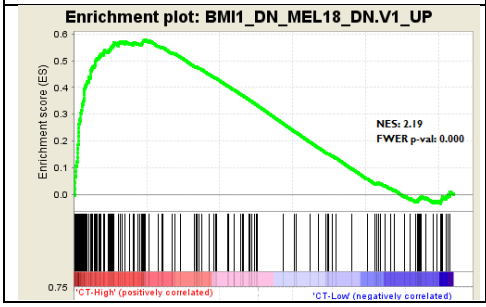
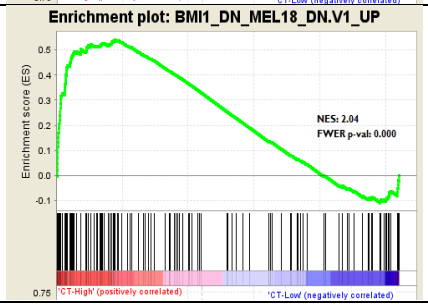
DDIT4	0.000204366	0.35551548	UP	GMIP	0.001111642	0.446264459	DOWN
AMPD3	0.00021365	0.348436506	DOWN	CXorf48	0.00112061	0.443048455	UP
TMEM164	0.000234626	0.360137696	UP	LOC100506934	0.001139576	0.443822509	UP
RANGAP1	0.000236752	0.343210924	DOWN	SEZ6L	0.001158288	0.444475829	UP
RNU5F-1	0.000238391	0.327398645	UP	AAMP	0.001161515	0.439254629	DOWN
GRID2IP	0.000297744	0.388466909	UP	LETM2	0.001170907	0.436480848	DOWN
FAM108A4P	0.000309949	0.385134182	UP	ACE2	0.001185515	0.435701938	UP
FAM83G	0.000371237	0.440320911	DOWN	LOC401022	0.001196965	0.433799961	UP
IQCD	0.000414568	0.470336972	UP	TOMM20	0.001243023	0.444321206	UP
KDM3A	0.000428303	0.465672375	UP	HEATR7A	0.001246462	0.439529375	DOWN
LANCL3	0.000458449	0.478510619	UP	B4GALT5	0.001247636	0.434077529	DOWN
KIAA0226	0.000468532	0.470226399	DOWN	HSD3B7	0.00130297	0.447364439	DOWN
ECHDC3	0.000473321	0.457438397	UP	LOC100506090	0.001319405	0.447124237	UP
C9orf41	0.000486464	0.453349523	DOWN	TCEAL6	0.001331342	0.445385068	DOWN
ZNF44	0.000491623	0.442359116	DOWN	MC4R	0.001391799	0.459716443	UP
GALNTL5	0.000497108	0.432384329	UP	CHST15	0.001401022	0.456978442	DOWN
CDC42EP2	0.000499335	0.420310797	DOWN	OVCA2	0.00140268	0.451870639	DOWN
ZNF322P1	0.000502863	0.410053271	UP	LRRIQ1	0.00140981	0.448629014	UP
PLEKHG5	0.000530095	0.419160741	DOWN	C20orf141	0.001418523	0.445963214	DOWN
A2MP1	0.00061238	0.469983504	UP	SLC4A11	0.001418551	0.440662864	DOWN
WAPAL	0.000628346	0.468458915	DOWN	SV2B	0.001422958	0.436831483	UP
FAM184A	0.000644123	0.466882156	UP	TNFRSF25	0.001429402	0.433707049	DOWN
BAI3	0.00065527	0.462124883	UP	ZNF813	0.001462199	0.43855886	UP
ATP6V1C2	0.000656261	0.450644042	UP	SHBG	0.001493564	0.442875694	DOWN
ERO1LB	0.000665316	0.445147891	UP	ALG10B	0.001537168	0.450683873	UP
TP63	0.000673646	0.439452738	UP	LOC439990	0.001560095	0.452323512	DOWN
LOC283194	0.000682284	0.434232248	UP	DNAJB12	0.00159745	0.458064318	DOWN
ZBTB11	0.000705668	0.438421607	UP	FLJ46257	0.00168324	0.477418167	UP
ZBTB46	0.000747141	0.453392917	DOWN	FCGR2B	0.001752444	0.491701803	UP
CLU	0.000788753	0.467766328	DOWN	HSD17B11	0.001756218	0.487518726	UP
RLF	0.000845233	0.490122384	UP	PPP1R7	0.00176544	0.484919897	DOWN
MON1A	0.000845437	0.479583468	DOWN	RG9MTD2	0.001766848	0.480251357	UP
IFITM2	0.000848584	0.47112681	DOWN	MIR548AN	0.001786161	0.480495792	UP
HDAC7	0.000854181	0.464354204	DOWN	AIMP1	0.001786701	0.475736371	UP
BRD2	0.000913415	0.486421381	UP	BEND5	0.001789312	0.471619145	UP
MIR922	0.000922401	0.481382625	DOWN	VANGL2	0.001798915	0.469408866	UP

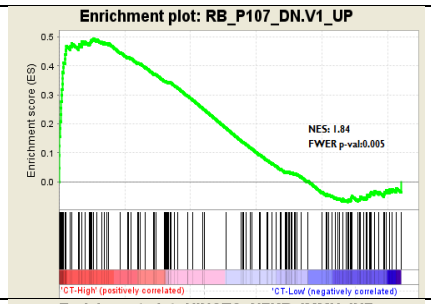
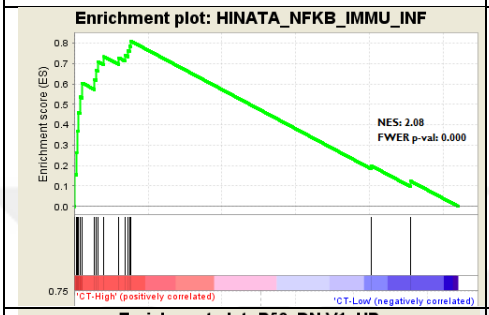
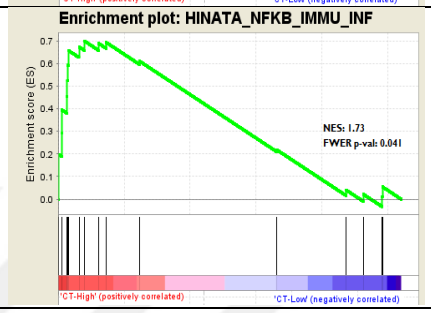
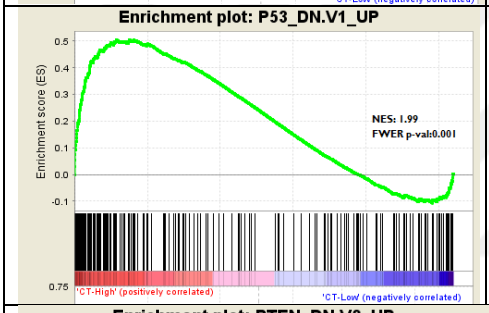
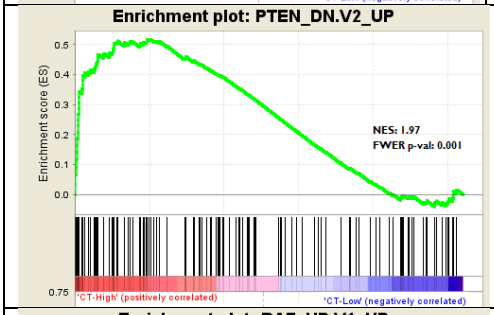
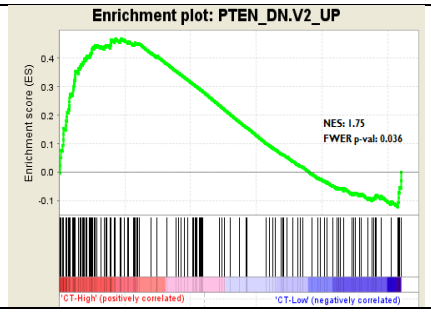
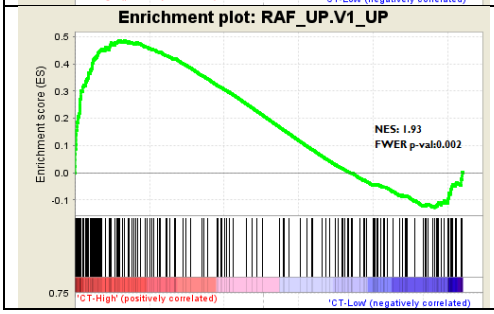
Supplementary Table 1.11 Differentially expressed transcripts between CT-High and CT-Low skin mesenchymal cancer cell lines. Top 100 genes were listed below.

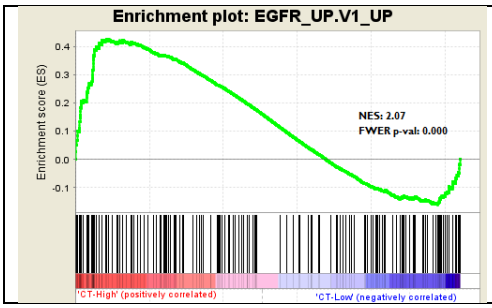
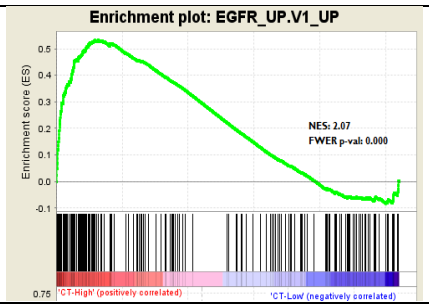
MELANOMA/MESENCHYMAL							
Gene Name	**p-value**	**Adjusted p-val**	**Direction**	**Gene Name**	**p-value**	**Adjusted p-val**	**Direction**
LOC100653084	6.53129E-11	1.70427E-06	UP	LRRK1	4.08E-05	0.020871	DOWN
LOC100288568	2.56168E-10	3.34222E-06	UP	ZDHHC19	4.09E-05	0.020526	UP
TPTE	6.9018E-10	6.00319E-06	UP	CT45A6	4.39E-05	0.021628	UP
LOC100291796	8.91879E-10	5.81817E-06	UP	TAL1	4.44E-05	0.021468	UP
LINC00221	6.4108E-09	3.34567E-05	UP	NFE2	4.68E-05	0.022211	UP
DSCR4	9.50306E-09	4.13288E-05	UP	C4orf39	5.13E-05	0.023909	UP
LOC100508797	1.18207E-08	4.40643E-05	UP	LOC100505490	5.21E-05	0.023856	UP
C22orf34	1.28702E-08	4.19793E-05	UP	SLCO1A2	5.3E-05	0.023855	UP
BAGE4	1.32508E-08	3.84184E-05	UP	ATP1B2	5.5E-05	0.024327	UP
BAGE3	1.34344E-08	3.50558E-05	UP	MAGEC1	5.53E-05	0.024036	UP
BAGE2	1.64937E-08	3.9126E-05	UP	RERG	5.55E-05	0.023752	UP
TAG	5.81796E-08	0.000126512	UP	LOC100507370	5.59E-05	0.023519	UP
PAGE2B	9.44419E-08	0.000189567	UP	WDR72	5.76E-05	0.02385	UP
LOC400643	1.5131E-07	0.000282019	UP	HSD17B3	6.3E-05	0.025687	UP
FLJ45974	3.64587E-07	0.000634235	UP	KCNH5	6.7E-05	0.026896	UP
MGC39584	7.95142E-07	0.001296777	UP	C3orf37	7.68E-05	0.030382	UP
CTAG2	8.37541E-07	0.001285577	UP	PGRMC1	8.33E-05	0.032429	DOWN
MAGEA12	8.74878E-07	0.001268281	UP	LOC100507599	8.85E-05	0.033978	UP
DSCR8	1.04651E-06	0.001437246	UP	LOC649395	9.73E-05	0.036798	UP
PAGE5	1.74915E-06	0.002282111	UP	TUT1	9.73E-05	0.036284	UP
XAGE1D	2.03176E-06	0.002524603	UP	MOCS3	9.84E-05	0.03616	UP
XAGE1E	2.05982E-06	0.00244314	UP	MYH1	9.89E-05	0.035853	UP
DHH	2.08717E-06	0.002367942	UP	MAGEB2	0.0001	0.035786	UP
CTAG1B	2.51088E-06	0.002729957	UP	ARF6	0.000102	0.03596	DOWN
MAGEA10	3.18497E-06	0.003324343	UP	FAM46D	0.000105	0.03644	UP
KPNA6	4.57753E-06	0.004594084	DOWN	RGS12	0.000105	0.036006	DOWN
ADAMTS20	4.68918E-06	0.00453183	UP	CSAG3	0.000107	0.036354	UP
MAGEA10	6.80972E-06	0.006346174	UP	LOC170425	0.000109	0.036315	UP
LOC100128737	7.12835E-06	0.006414041	UP	RBBP4	0.000111	0.03661	DOWN
LOC100507559	9.82418E-06	0.008545075	UP	LOC100147773	0.000112	0.036557	UP
LOC100505874	1.09904E-05	0.00925111	UP	LOC386758	0.000114	0.03673	UP
CSMD1	1.34703E-05	0.01098422	UP	OR8A1	0.000119	0.037848	UP
TNNI3	1.37878E-05	0.010902411	UP	METTL7B	0.000121	0.03797	UP
NAA11	1.45828E-05	0.011191874	UP	KC6	0.000126	0.039051	UP
MAGEA11	1.51158E-05	0.011269467	UP	ZNF37A	0.000135	0.041474	DOWN

CSAG1	1.53126E-05	0.011099094	UP	MYH13	0.000138	0.041725	UP
MAGEA3	1.71042E-05	0.012062593	UP	CTAG1A	0.000141	0.042235	UP
FLJ36000	1.89811E-05	0.013034025	UP	LOC100509302	0.000161	0.047652	UP
MAGEA5	1.9525E-05	0.013063712	UP	CT45A1	0.000168	0.049149	UP
TMEM57	2.32677E-05	0.01517866	DOWN	GABRG2	0.000171	0.049451	UP
LOC100506433	2.34294E-05	0.014911356	UP	GPRC5D	0.000172	0.049367	UP
OXGR1	2.41898E-05	0.0150288	UP	OR8G5	0.000172	0.048925	UP
SYNC	2.44196E-05	0.014818746	DOWN	DUX4L9	0.000179	0.050155	UP
SLCO1B1	2.4892E-05	0.014762094	UP	LOC100009676	0.00018	0.049896	UP
LOC100289097	3.06707E-05	0.01778493	UP	CCDC71	0.000186	0.051079	UP
GNN	3.40195E-05	0.019297957	UP	JAK1	0.000187	0.050942	DOWN
MAGEA2	3.41727E-05	0.018972366	UP	FLJ43315	0.000188	0.050552	UP
ZNF204P	3.42676E-05	0.018628709	UP	CSAG2	0.000189	0.050413	UP
MYH8	3.70439E-05	0.01972699	UP	SELK	0.000191	0.050315	UP
LOC100508631	3.84113E-05	0.020046066	UP	CT45A4	0.0002	0.052257	UP

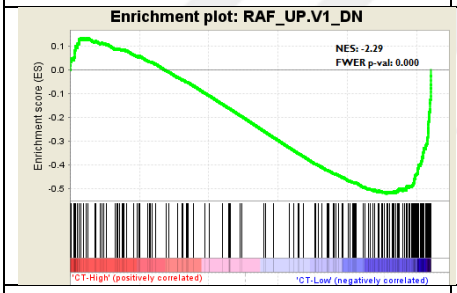
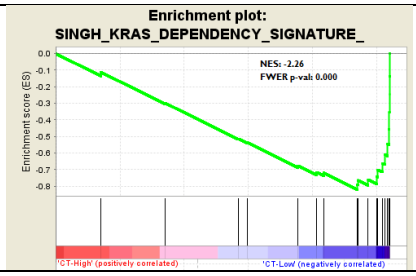
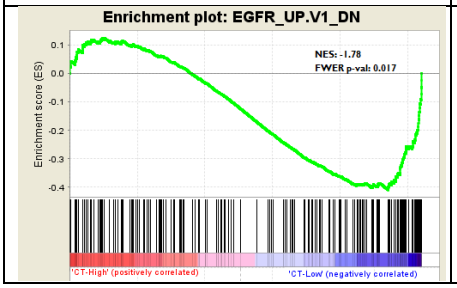
Supplementary Table 1.12: Gene set enrichments in CT-High breast cancer cells. GSEA is performed with CT-High and CT-Low subgroups of breast cancer cell lines with respect to epithelial or mesenchymal phenotype.

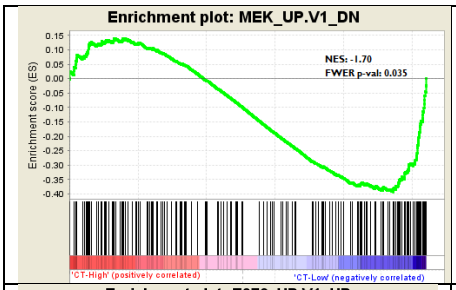
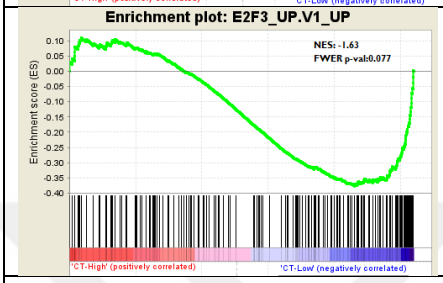
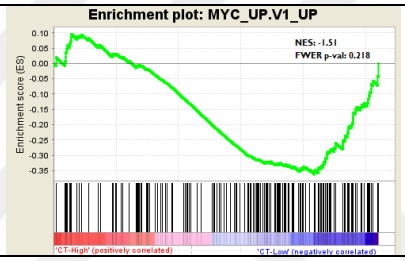
Breast/Epithelial	Breast/Mesenchymal	Gene Set Description
		Genes regulated by Hippo Pathway, related with CSCs in breast
		Genes upregulated in medullablastoma cells following knockdown of transcriptional repressor proteins Bmi-1 and Mel-18

		Genes upregulated in primary keratinocytes from RB1 and RBL1 skin specific knockout mice
		Inflammatory and immune genes induced by NF-kappaB in primary keratinocytes and fibroblasts
		Genes upregulated in cell lines with mutated TP53
		Genes upregulated in HCT116 cells following knockdown of PTEN
		Genes upregulated in MCF-7 cells stably over-expressing constitutively active RAF1 gene

		Genes upregulated in MCF-7 cells engineered to express ligand eligible EGFR.

Supplementary Table 1.13: Gene set enrichments in CT-Low breast cancer cells. GSEA is performed with CT-High and CT-Low subgroups of breast cancer cell lines with respect to epithelial or mesenchymal phenotype.

Breast/Epithelial	Breast/Mesenchymal	Gene Set Description
		Genes downregulated in MCF-7 cells stably over-expressing constitutively active RAF1 gene
		Genes defining KRAS dependency signature
		Genes downregulated in MCF-7 cells engineered to express ligand eligible EGFR.

		Genes downregulated in MCF-7 cells stably over-expressing constitutively active MAP2K1 gene
		Genes upregulated in primary breast cancer cell culture over-expressing E2F3 gene
		Genes upregulated in primary breast cancer cell culture over-expressing MYC gene