ROLE OF OVARIAN CANCER-INITIATING CELLS IN HIGH-GRADE SEROUS OVARIAN CARCINOGENESIS

Rohit Jadhav

Submitted to the faculty of the School of Informatics,

in partial fulfillment of the requirements

for the degree of

Master of Science in Bioinformatics,

Indiana University

December 2011

Accepted by the Faculty of Indiana University,

in partial fulfillment of the requirements for the degree of Master of Science

in Bioinformatics

Master's Thesis

Committee

_____

Yaoqi Zhou, Ph.D., Chair

_____

Kenneth P. Nephew, Ph.D.

_____

Lang Li, Ph.D.

_____

Changyu Shen, Ph.D.

_____

Narayanan B Perumal, Ph.D.

Dedicated to my Family

Contents

TABLE OF FIGURES

ACKNOWLEDGEMENTS

I would like to take the opportunity to acknowledge some of the people who made my graduate study a memorable experience and made this thesis possible. Words are really inadequate to express the deep sense of my gratitude and regards towards my advisors Dr. Kenneth Nephew, Dr. Lang Li and Dr. Changyu Shen. Their valuable advice, motivation, close supervision, feedback, and encouragement inspired me to complete this work successfully.

I would like to thank Dr. Curt Balch for his time, encouragement, insightful comments, and critical feedback. I greatly appreciate his efforts in reviewing my work. Inspite of his tight work schedules, he has helped me unconditionally. I would also like to thank Dr. Narayanan B. Perumal, for his constant support and guidance throughout the program. I owe my sincere thanks to Indiana University for providing the prestigious Dean's award during my Master's program. I would like to extend my sincere thanks to Dr. Yaoqi Zhou and the faculty and staff of the department of Bioinformatics for their continuous support throughout the Masters program.

Above all, I express my deep gratitude and affection to my parents whose persistent love, warmth and cooperation was an encouragement for successful completion of my thesis. I would like to extend my loving thanks to Ms. Amruta Deshpande for her support throughout my Master's program. I owe my sincere thanks to Ms. Anagha Chitale for her guidance and support during my stay in United States. I would also like to thank my friends, Yogesh, Shreyas, Sarang, Prasad, Sunil, Gaurav and Abhishek for their support and all the fun we have had in the last two years.

ABSTRACT

A subpopulation of tumor cells known as ovarian cancer initiating cells (OCICs) have been shown to be the cells that propagate the tumor phenotype in ovarian cancer. Studies have showed that a very small population (100) of these cells is sufficient to induce a tumor phenotype; while a large quantity of tumor cells ($5 \text{ X } 10^5$) are required to induce such a phenotype. In this study we studied the functional changes in genes expressed in the OCIC phenotype which were important for such efficient propagation of cancers. To enable this analysis, we generated mRNA expression and DNA methylation profiles of OCICs and compared them with those of tumor and normal ovarian surface epithelial cells. We identified four pathways which regulated most of the observed changes and were predicted to be important factors in distinguishing the OCICs from tumors and normal cells. The gene signatures for these pathways were analyzed by unsupervised clustering in order to determine the similarities of OCICs with respect to tumor and normal samples. We further believed that the OCICs can be used as indicators towards the genesis and progression of early events in the ovarian cancers. In light of this, we considered two hypotheses which are currently addressing the genesis of ovarian cancer. The first hypothesis proposed ovarian surface epithelial cells to be cells of origin of the ovarian cancer while the other proposed the fallopian tube cells to be contributing the cell of origin for these cancers. It is also believed that these two cells can be reciprocal cells of origin for the cancer phenotype. In order to test these hypotheses, we integrated the in-house dataset with a public domain fallopian tube gene expression data. The integration of the results obtained from these analyses provided better understanding of the early events in ovarian carcinogenesis.

Chapter 1 INTRODUCTION

Ovarian Cancer

Ovarian cancer is the most lethal gynecological neoplasm, with more than 100,000 cancer deaths among women worldwide each year, making it as the fifth leading cause of cancer deaths among woman (Bowtell, 2010). Moreover, the poor 5-year survival (30-40%) results from the high percentage of cases diagnosed at an advanced stage and is largely due to the fact that most of them are inoperable when first discovered and respond poorly to therapy(Auersperg, Wong, Choi, Kang, & Leung, 2001; Landen, Birrer, & Sood, 2008).

Ovarian cancers according to National Cancer Institute (NCI) are categorized into 4 stages depending on how far the disease has spread. Stage I means that the cancer has been expressed on the ovary only, Stage II the cancer has spread to the pelvis, at Stage III the cancer has spread to the abdomen and finally Stage IV is when the cancer has spread outside the abdomen. Depending at which stage the disease is diagnosed the survival rate will improve dramatically. If the diagnosis is made in the early stages of the disease (Stage I) and the patient gets treatment before the cancer spreads outside the ovary, then a 5 year survival rate of 94% can be achieved. The early physical symptoms of ovarian cancer in Stage I and II are common to other benign diseases such as low back pain, pelvic discomfort, loss of appetite and others. These are mild symptoms not specific to ovarian cancer which makes the early detection harder. The need for a screening test that will detect ovarian cancer at the early stages of the disease is clear from these facts. The disease prevalence in post-menopause women is 40 in 100000 which leads to strict

specificity requirements from a screening test to avoid having too many false positives and thus fewer women going through unnecessary operations and treatment.

Classes of Ovarian Cancer

Ovarian cancers have been mainly classified into 2 types: Type 1 and Type 2.

Type 1 tumors include endo-metrioid, mucinous and low grade serous cancers which are usually low grade and slowly developing. They show a high frequency of Ras pathway mutations and generally lack TP53 mutations. They are shown to have a relatively normal karyotype and are often poorly responsive to platinum-based therapy.

Type 2 tumors are rapidly progressing high-grade serous carcinomas without a well defined premalignant lesion. They are strongly associated with TP53 mutations and rarely have Ras mutations. This subtype accounts for about 60% to 80% of the cancer cases and is the most aggressive. Among these cases, less than 25% are detected at an early stage (stages I and II). As Type 2 tumors account for the bulk of high-grade serous ovarian cancer, the terms Type 2 and high-grade are largely equivalent(Bowtell, 2010; Levanon, Crum, & Drapkin, 2008). This thesis focuses on the High-grade or Type-2 serous ovarian cancer (HG-SOC).

Diagnosis of Ovarian Cancer

Early diagnosis is of utmost importance for the successful treatment, and survival from, ovarian cancer. However, in most of the cases, ovarian cancer is diagnosed at an advanced stage when the probability of survival is very less. The prospect of early diagnosis is also impeded by several other aspects of the disease. First, in contrast to many other malignant diseases the cell of origin of ovarian cancer and the early events of ovarian carcinogenesis, have not been clearly identified (Dubeau, 2008). Second, current

methods of examination of the ovaries can only be made by invasive diagnostic means, that is, via a pelvic examination, including the palpation of the ovaries, and with varying success, by ultrasound. In addition, proteins such as CA125 and other tumor markers have proved inefficient due to inherent low sensitivity. Methylation markers, however, potentially represent a potent new tool for early detection. Aberrant methylation is thought to be one of the earliest molecular changes in carcinogenesis (A. P. Bird & Wolffe, 1999), and as such can be applied for the detection of early-stage or potentially premalignant disease.

Role of Cancer Stem Cells

It has been suggested that tumor re-growth, as well as chemotherapy resistance and metastasis, are dependent on a small sub-population of cancer cells within the tumor that are thought to represent cancer stem cells (CSCs). A defining hallmark of stem cells, in both normal and malignant tissue, is the ability to self-renew but simultaneously give rise to daughter cells that are committed to differentiation into phenotypes that often cross lineages. To achieve this, stem cells can undergo an asymmetric cell division whereby they segregate cell fate determinants into only one of the two daughter cells (Knoblich, 2008). In adult mammals, stem cells have been characterized for a number of tissues, including the blood system, central nervous system, muscle, colon, breast, and bone/cartilage.

For cancer stem cells in solid tumors, much can be learned from studies of hematopoietic stem cells (HSC) for which it has been shown that the transplantation of a single cell into a myeloablated recipient can reconstitute the entire blood system. This definitive HSC gives rise to a hierarchy of pluripotent progenitors that become

progressively restricted in their differentiation potential. Leukemia is thought to originate either from HSCs that acquired genetic or epigenetic changes and became partially differentiated and tumorigenic, or from progenitors that acquired the capacity to self-renew (Passegue, Jamieson, Ailles, & Weissman, 2003). The existence of stem cells for certain types of leukemia is strongly supported by lentiviral tagging of human acute myelogenous leukemia cells and the observation of individual clones present in NOD–SCID mice after serial transplantation of the tagged cells (Hope, Jin, & Dick, 2004). Studies of leukemia stem cells also indicated great phenotypic plasticity depending on the stage of tumor growth, tumor microenvironment, and external factors such as stress created by radio-or chemotherapy (Passegue, et al., 2003).The presence of CSCs in solid tumors has been proposed for human cancers including breast (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003), brain (Galli et al., 2004), colon (O'Brien, Pollett, Gallinger, & Dick, 2007), head and neck (Prince et al., 2007), pancreatic (Ophorst et al., 2007), prostate (Collins, Berry, Hyde, Stower, & Maitland, 2005), ovarian (Baba et al., 2009; Bapat, Mali, Koppikar, & Kurrey, 2005; Zhang et al., 2008), and skin cancer (Fang et al., 2005). The characteristics of solid tumor stem cells have been defined as ‘‘a small subset of cancer cells within a cancer that constitute a reservoir of self-sustaining cells with the exclusive ability to self-renew and to cause the heterogeneous lineages of cancer cells that comprise the tumor’’ (Clarke et al., 2006). Several cell surface markers, including CD24, CD44, CD133, CD166, EpCAM, or dye efflux assays have been used to sort populations of putative cancer stem cells from primary tumor cultures or cell suspensions obtained from tumor biopsies. After transplantation into immunodeficient mice, tumors form from several hundred marker-positive cells, whereas for marker

negative cells orders of magnitudes higher numbers are needed to achieve the same frequency of tumor formation (Gupta, Chaffer, & Weinberg, 2009). Recently, using an improved xenotransplantation technique, a study with human melanoma cells showed tumor formation after inoculation of one tumor cell (Quintana et al., 2008) and thereby an important step towards the proof of CSC existence was made (Strauss et al., 2011).

Epigenetics

The epigenome provides a mechanism of cellular memory and is key in regulating and deciphering stages of normal and abnormal cellular development, including phases of growth, differentiation, senescence, aging and immortalization during carcinogenesis (Feinberg & Tycko, 2004). The components of the epigenome, DNA methylation, post-translational histone and other protein modifications, nucleosome positioning and noncoding RNAs (specifically microRNA [miR] expression), act in concert to exert their cellular effects (Sharma, Kelly, & Jones, 2010).

Epigenetics is defined as the study of heritable changes in gene regulation without a change in the DNA sequence or the sequence of proteins associated with DNA. Histone modifications and DNA methylations are the most studied epigenetic changes in cancer biology. DNA methylation primarily occurs in the CpG dinucleotides and is often altered in the cancer cells. Both hypomethylation and hypermethylation contribute to tumorigenesis. Gene-specific hypermethylation is observed in many different types of cancers and is often seen responsible for the silencing of tumor suppressor genes (Jones & Laird, 1999). On the other hand, hypomethylation is responsible for the development of genomic instability mostly by the activation of oncogenes or loss of imprinting which

is frequently observed in cancer (Feinberg & Tycko, 2004; Liu, Wylie, Andrews, & Tollefsbol, 2003).

Histone modifications are also important in cancer biology and are responsible for major changes in chromatin structure affecting the accessibility of DNA to the transcription factors which regulate gene expression. Mainly, histone acetylation has been known to increase the gene expression affecting from the uncoiling of the chromatin thus making the DNA accessible to the transcription factors. On the other hand histone deacetylation leads to coiling of the chromatin structure thus preventing gene expression as the DNA is rendered inaccessible to the transcription factors(Verdone, Caserta, & Di Mauro, 2005). Other types of histone modifications like histone methylation show varied effects and all these alterations together seen in cancer have major role in the initiation of cancer and its progression to potentially malignant cells (Lehrmann, Pritchard, & Harel-Bellan, 2002). However, for this study we focus mainly on DNA methylation and its role in cancer which will be discussed in detail further.

DNA Methylation

DNA methylation is an addition of a methyl group to the C-5 position of a cytosine residue, usually in the context of CpG dinucleotides (A. Bird, 2002; Das & Singal, 2004). As the only known mammalian modification of DNA itself (Robertson, 2005), DNA methylation is essential for parental imprinting, embryonic development, gene regulation, and chromosomal stability (Bartolomei & Tilghman, 1997; Herman, 1999; Jaenisch, Beard, Lee, Marahrens, & Panning, 1998; Li, Bestor, & Jaenisch, 1992). The link between DNA methylation and cancer dates back to 1983, when various investigators found that cancer cells possess an inverse methylation pattern to that of

normal cells (Feinberg & Tycko, 2004). In normal cells, the genome is overall hypermethylated, especially in repetitive genomic elements (A. Bird, 2002), while small stretches of CG-rich promoter regions, "CpG islands" (CGI), remain largely unmethylated (Baylin, 2005). In tumor cells, however, global DNA hypomethylation is observed while CGI, by contrast, acquire hypermethylation (Jones & Baylin, 2002). This aberrant CGI hypermethylation typically leads to silencing of associated coding sequences, providing an alternative to mutation or deletion for the inactivation of tumor suppressor genes (Luczak & Jagodzinski, 2006; Nephew & Huang, 2003). DNA methylation is catalyzed by a group of DNA methyltransferases (DNMTs): DNMT1, DNMT2, DNMT3A and DNMT3B (Pradhan & Esteve, 2003). DNMT1 primarily maintains pre-existing genomic methylation patterns, methylating hemimethylated DNA during DNA replication (Bestor, Laudano, Mattaliano, & Ingram, 1988). DNMT3A and DNMT3B, by contrast, are mainly responsible for *de novo* methylation, methylating both unmethylated and hemimethylated DNA with equal efficiency (Okano, Takebayashi, Okumura, & Li, 1999; Xie et al., 1999). The function of DNMT2 is little known, but has recently been reported to be a methyltransferase that uses tRNA as a substrate (Goll et al., 2006).

Despite the role of promoter CGI methylation in tumorigenesis, recent studies indicate that it may also play a prominent role in the complexity of cancer drug resistance (Fojo & Bates, 2003). One extensively studied example is the DNA mismatch repair enzyme human Mut-L homologue (hMLH1). As a key component in the mismatch repair system, hMLH1 recognizes drug-DNA adducts and initiates drug-induced apoptosis (Vaisman et al., 1998). Epigenetic silencing of *hMLH1* by methylation leads to

diminished apoptosis response to bulky drug-DNA adducts, and therefore is one contributing factor to drug-resistance development (Balch et al., 2005; Siddik, 2003). Besides *hMLH1*, several other chemotherapy responsive genes are also deregulated during the acquisition of chemoresistance (Fraser et al., 2003; Grossman & Altieri, 2001; Pommier, Sordet, Antony, Hayward, & Kohn, 2004).

Recently, the technological advancements have enabled quantitative assessments of thousands of loci for DNA methylation status, thus providing an opportunity to study the epigenetic signature of a cell. This rich source of information can be used to better understand the steps of carcinogenesis and their cells of origin. Further it can be used to identify appropriate model systems for further studies and also to identify potential biomarkers detection, classification and monitoring or the disease (Houshdaran et al., 2010).

Techniques in DNA methylation studies

DNA methylation has been historically studied in a locus-targeted manner. However, with the advent of highthroughput platforms, large-scale structure of genomic methylation patterns is available through genome-wide scans. Two high-throughput platforms that have been popularly used include the Illumina Infinium Human Methylation27 array and the Illumina GoldenGate array. Both arrays are based on genotyping bisulfite (BS)-converted DNA. DNA samples are treated with a methylation kit that converts unmethylated cytosines to uracils, whereas methylated cytosines are protected and remain cytosine. Therefore, whether the base at a given locus is converted or not provides information on its original methylation status. The results of the array, the methylation status of the interrogated CpG site is a sequence of β-values, one for each

locus, calculated as the average of approximately 30 replicates (with approximately 30 beads per site per sample) of the quantity max(M, 0)/(max(U,0)1max(M,0)1100). Here U is the fluorescent signal from an unmethylated allele on a single bead, M is that from a methylated allele. A maximum between signal intensity and 0 is chosen to compensate for negative signals due to background subtraction. The constant 100 is to regularize b-values when both M and U values are small (Bibikova et al., 2006).

As explained earlier, β-value ranges continuously from 0 (unmethylated) to 1 (completely methylated) and reflects the methylation level of each CpG site. It can be thought of as a ratio of the methylated (or unmethylated) probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities). According to the notation used by Illumina methylation assay, Beta-value for an i^{th} interrogated CpG site is defined as:

$$Beta_i = \frac{\max(y_{i,methy}, 0)}{\max(y_{i,unmethy}, 0) + \max(y_{i,methy}, 0) + \alpha}$$

where $y_{i,menty}$ and $y_{i,unmenty}$ are the intensities measured by the i^{th} methylated and unmethylated probes, respectively. To avoid negative values after background adjustment, any negative values will be reset to 0. Illumina recommends adding a constant (stated earlier) offset α (by default, α = 100) to the denominator to regularize Beta value when both methylated and unmethylated probe intensities are low. The Beta-value statistic results in a number between 0 and 1, or 0 and 100%. Under ideal conditions, a value of zero indicates that all copies of the CpG site in the sample were completely unmethylated (no methylated molecules were measured) and a value of one indicates that every copy of the site was methylated (Du et al., 2010).

Recently, tremendous amounts of DNA methylation data have been generated from high- throughput DNA methylation platforms. Currently, to select differentially methylated loci, researchers mainly apply either parametric methods such as regression-based methods or t-test or nonparametric methods such as rank sum test. Many studies have shown that β-values generated by BeadStudio with Illumina arrays usually have a heavy tail close to zero which represents unmethylated and a bump close to one which represents completely methylated. One such study by Shuang Wang (Wang, 2011) studied these patterns and proposed statistical approaches for analyzing the methylation data. This study used a β-value threshold of 0.5 to differentiate between unmethylated and completely methylated loci, with those above β-value of 0.5 considered as completely methylated and the loci below as unmehtylated. They further proposed a method to select differentially methylated loci between ovarian cancer cases and age-matched healthy controls using Illumina Infinium Human Methylation27 Beadchip (Teschendorff et al., 2010) and identified some methylation loci that are missed by the existing method.

Chapter 2 BACKGROUND

Predictions for the Cell of origin in Ovarian Cancer

It has been thought that serous ovarian cancer (SOC) arise from ovarian surface epithelium or intra-ovarian inclusion cysts. This theory has recently been challenged by the identification of tubal intraepithelial carcinoma (TIC) in woman with BRCA1 or BRCA2 mutations found in the distal end of fallopian tube which is proposed as probable precursors of advanced HG-SOC. The gene signatures of fallopian tube epithelium and HG-SOC are found to be similar and the coexistence of TIC with similar TP53 mutations support the theory that fallopian tube is an important site for the initiation of high-grade serous ovarian cancer (HG-SOC). However, the contribution of fallopian sites to the genesis of HG-SOC is unclear at present. The wide-spread growth of tumor tissue in advanced cancers usually obscures the primary site of origin. Even though the primary tumor seems to originate from the ovary, the theory suggests that the fallopian tube may have provided the originating cell supposedly through entrapment in the ovary (endosalpingiosis)(Bowtell, 2010).

Ovarian Cancer Initiating Cells

In a previous study, Zhang et. al isolated and characterized ovarian cancer initiating cells (OCICs) using primary human ovarian tumors. These OCICs were seen to be fully capable of reestablishing their original tumor phenotype in vivo. These cells could be isolated using antibodies against CD44 and CD117. These could propagate tumors in animals and fulfilled the current cancer stem cell criteria including self-renewal, small minority of total tumor population, reproducible tumor phenotype, multipotent differentiation into nontumorigenic cells and having distinct cell surface antigenic phenotype, permitting consistent isolation.

This study further investigated the tumorigenic potential of $CD44^{+}CD117^{+}$ cells (Cells having these markers) compared to that of $CD44^{-}CD117^{-}$ (Cells without the markers). They observed that as few as 100 cells with the markers were required to result tumor formation in mice, whereas 5 X 10^{5} cells without the markers were required to develop the tumor.

These cells have been thought to provide a better insight into the genesis of the ovarian cancer. The speculated origin for the OCICs is the ovarian surface epithelium (OSE). This is based on the fact that like OCICs, OSE are also known to express CD44 and CD117 surface markers. However, based on the histology and gene expression patterns, the ovarian cancers are hypothesized to be derived from fallopian Tube epithelial cells. This is because the ovarian tumors are known to become epithelial during tumor progression. On the other hand, OSE are known to retain capacity to undergo epithelial-to-mesenchymal transition required for postovulatory repair mechanism. It is thought that the comprehensive studies of epigenetic patterns (Histone modifications & DNA Methylation) in OCICs will lead to better understanding of the early events in ovarian cancer(Zhang, et al., 2008).

Epithelial – Mesenchymal Transition

The changes observed in the cell phenotypes between epithelial and mesenchymal states are defined as epithelial – mesenchymal (EMT) and mesenchymal – epithelial (MET) transitions. They have been known to play a key role in the process of embryonic development and also have recently known to be important for the pathogenesis of various cancers. EMT is a complex program through which an epithelial cell loses its characteristic differentiated phenotype by the accompanied loss of functions like cell-cell

adhesion, lack of motility, planar and apical-basal polarity. It in-turn acquires mesenchymal characteristics like motility, invasiveness and increased resistance to apoptosis.

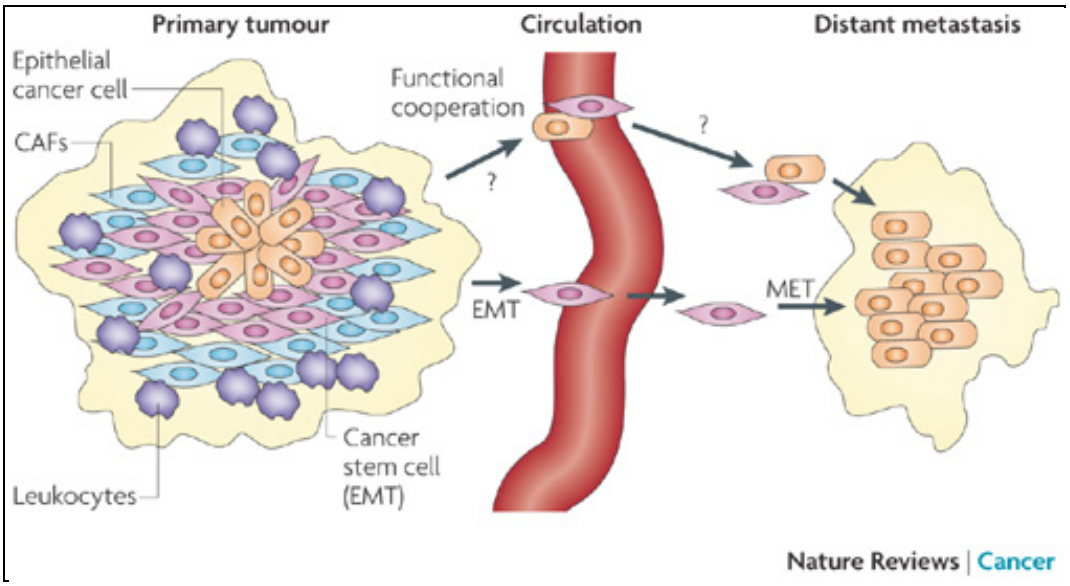

Figure 1: Epithelial-to-Mesenchymal Transition and the Mesenchymal-to-Epithelial Transition in Ovarian Cancer (Polyak & Weinberg, 2009)

EMT has been known to be the important factor in the metastatic dissemination, invasiveness and resistance to therapy. On the other hand MET is the reversal of EMT and is observed after the dissemination and is responsible for the formation of metastases in remote locations. However, unlike embryogenesis, these programs in cancer are mostly irreversible events as they are caused due to somatic mutations (Polyak & Weinberg, 2009).

TGF-β Signature

The Transforming growth factor-β (TGF- β) family of cytokines has been known to induce the process of EMT during the embryonic development, various diseases and cancer pathogenesis. However, recently TG1F- β has been shown to play a critical role in the regulation of breast cancer stem cell phenotypes. It has known to induce EMTs

through multiple signaling mechanisms which include phosphorylation by the receptors of SMAD transcription factors and proteins present in the cytoplasm, regulating cell polarity and formation of tight junctions. In the mammary epithelial cells, it is known that TGF- β type II receptor directly phosphorylates SMAD2, SMAD3 and protein regulation cell polarity PAR6A. The phosphorylation of this protein (PAR6A) reduces the apical-basal polarity and consequently the tight junctions between adjacent epithelial cells are dissolved. TGF- β is also known to affect the activities of various other EMT-inducing pathways like Notch, Wnt and integrin signaling which act together to induce EMT programs (Polyak & Weinberg, 2009).

DNA Methylation Studies in Ovarian Cancer

There have been various molecular approaches to identify tumor markers for the diagnosis and prognosis of ovarian cancer including complementary DNA microarrays. DNA methylation studies have been mostly investigating the methylation patterns based on supervised and unsupervised cluster analysis for identifications of different phenotypes and genes specific to them. Houshdaran et. al studied the methylation profiles of ovarian epithelial tumors and ovarian cell lines. They used Illumina GoldenGate platform with 1505 CpG sites representing 808 genes for studying 15 ovarian cell lines and 27 primary tumors. They considered measurements with a detection p-value less than 0.05 to have signal intensity significantly above background. They masked the data points with a detection p-value greater than 0.05 as “NA” considered them to be representing the beta values with non-significant detection of DNA methylation compared to background. Further, to identify the loci with significantly different levels of DNA methylation in cell lines compared to tumors, they restricted the analysis to 1,110

CpG sites whose beta-values were variable across samples. For this they considered those CpG sites which had minimum beta-value greater than 0.2, ratio of maximum beta-value to minimum beta-value greater than or equal to 2 and, 1 or fewer missing beta-values. They further performed two-sampled t-test for the 15 cell lines with 27 tumors and adjusting the p-values from the t-test for multiple comparisons by controlling the FDR. In the process, they identified 489 CpG sites representing 337 genes that varied significantly in the DNA-methylation at FDR less than 1%. They further identified that most of these i.e. 445 of 489 were highly methylated in cell lines compared to tumors. From this study they indicated the difference in DNA methylation profiles between ovarian cancers cell lines and tumors this emphasizing the cautiousness required in using the cell lines as models for tumors in further molecular studies. They also identified the different methylation patterns in the histological subtypes of ovarian cancer viz. Serous, Endometroid and Clear Cell carcinomas(Houshdaran, et al., 2010).

Another study by Michaelson-Cohen et. al used a genome-wide approach for identifying de novo methylated genes in epithelial ovarian cancer. They used methyl-DNA Immunoprecipitation and CpG Island microarrays for studying methylation patterns for 4 epithelial ovarian serous carcinoma and 2 normal ovarian tissue samples. After comparing these tissue types they identified 2583 islands that were constitutively methylated in most tissue types and 2484 (96%) of them were also methylated in the normal ovary. They further identified 16,962 unmethylated CpG Islands in normal ovary and found that 376 (2.16%) of these were methylated in ovarian cancer thus stating them as de novo methylated. Also, out of the 6,498 CpG islands that were methylated in normal ovary, 582 (9%) underwent demethylation in ovarian cancer. This implies that

these 582 islands were previously methylated in normal and underwent demethylation in cancer. However this study emphasized an unclear biological significance of the later group in tumorigenesis. It was also observed that the large number of normally methylated islands on the inactivated X chromosome underwent demethylation in ovarian cancer. This study thus provided a new perspective on methylation in ovarian cancer in a genome-wide approach. It also illustrated the way genes were silenced by methylation of CpG island and that the silenced genes played significant roles in the cell differentiation and functioning thus indicating potential biomarkers for diagnosis, prognosis and treatment of cancer(Michaelson-Cohen et al., 2011).

The Cancer Genome Atlas (TCGA) is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing. TCGA is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), two of the 27 Institutes and Centers of the National Institutes of Health, U.S. Department of Health and Human Services. TCGA undertakes comprehensive genomic characterization and analysis of each cancer and makes the data freely available to the cancer community through its data portal. It examine up to 500 samples for each tumor type thus providing statistical power to generate comprehensive genomic profiles of each cancer required for identification the best targets for drug development. For the entire sample types, both the normal and cancerous tissues are collected which allows identification of genomic changes that play roles in development of that particular cancer. Ovarian serous cystadenocarcinoma is the cancer being studied by TCGA and is a type of epithelial ovarian cancer which accounts for about 90 percent

of all ovarian cancers. The study mainly aims to determine gene expression patterns which are linked to differences in patient survival, to establish whether certain gene changes are linked to response to therapy and define the role of copy number variation (abnormal duplications or deletions) in cancer development. One of the findings of this study report is the identification of 4 subtypes of high grade serous ovarian cancer using 583 tumor samples. These subtypes include immunoreactive, differentiated, proliferative and mesenchymal. These were obtained by performing an integrated analysis of mRNA, miRNA expression and DNA methylation data from TCGA("Integrated genomic analyses of ovarian carcinoma," 2011).

Significance of studying OCICs

Ovarian cancer initiating cells have yet to be studied for specific changes in the expression patterns of genes and their corresponding methylation patterns. The study of changes in the pathways regulated by these genes will give us an insight in the progression and proliferation of ovarian cancer. Also these cells can be used as models for the studies on the genesis of ovarian cancer and also for testing the various hypotheses concerning the cell of origin. The inclusion of these cells for the comparison studies of tumors with fallopian epithelial cells and ovarian surface epithelial cells can help in determining the earlier events in the development of ovarian cancer.

Chapter 3 Hypotheses

Based on this information we formulated two hypotheses which will be tested by the studies in this thesis

1. Comparison of ovarian cancer initiating cells (“OCIC”) gene expression and DNA methylation profile to normal ovarian surface epithelium (“nOSE”) and bulk population tumor cells (“tumor”) can provide insight into specific pathways/processes involved in the initiation and progression of ovarian cancer (HG-SOC).
2. Comparison of gene expression signature concordances of ovarian cancer-initiating cells (OCICs) and various fallopian cancers/tissue and ovarian cancer/tissue can provide insight into the cell(s)-of-origin of high-grade serous ovarian cancer (HG-SOC).

Chapter 4 METHODS

Approach

In order to evaluate the proposed hypothesis, we formulated specific workflows for each separately. As these analyses required preprocessing of the raw datasets, they had to undergo various rigorous quality control steps which will be explained in the future sections. For the comparison of ovarian cancer initiating cells with tumors and normal cells, we developed a systematic workflow shown in Figure 1. Here we analyzed differentially expressed and methylated genes for similarly altered gene ontology pathways and also verified them by performing gene set enrichment analysis (GSEA) in case of expression datasets.

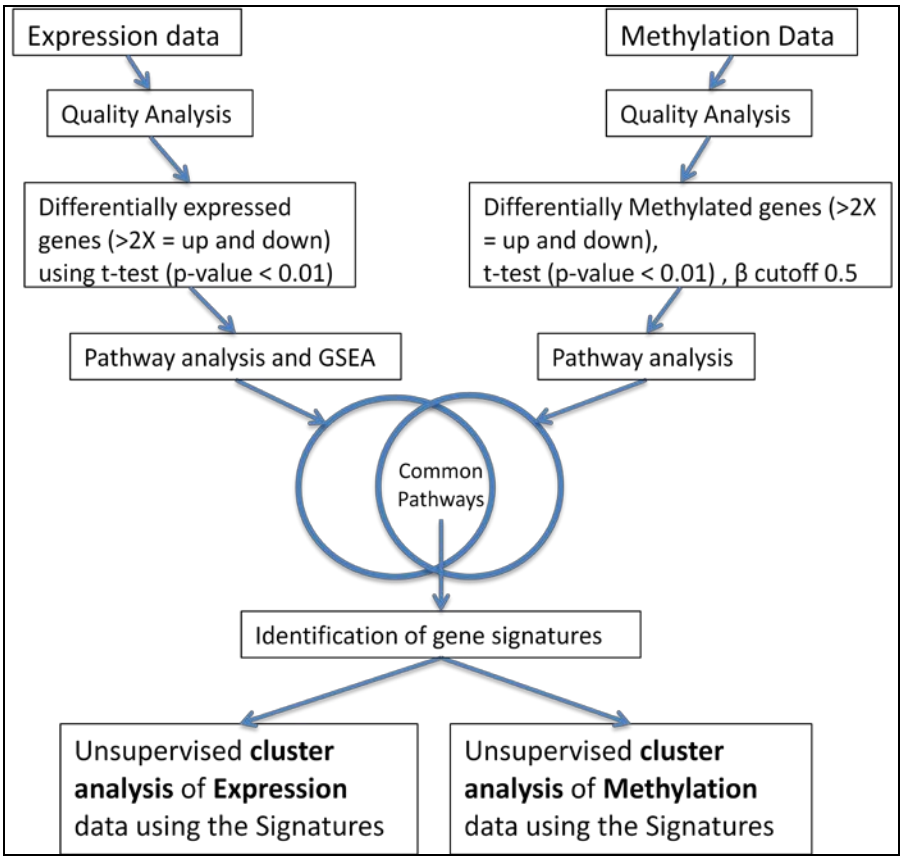

Figure 2: Analysis of Ovarian Cancer Initiating Cells using the gene expression and DNA methylation Datasets.

After determining the similarly altered pathways between these datasets, we set out to identify relevant gene signatures which will be efficient in indicating the difference in the three phenotypes. Using the information on these pathways and combining them with information from the literature, we identified these gene signatures. Further, we performed unsupervised cluster analysis for the two datasets separately for the genes in the signatures to get an indication of the differences in the signature patterns among the three phenotypes.

For the comparisons of ovarian cancer initiating cells with the fallopian tube datasets, we employed a separate workflow. This was mainly due to the fact that the fallopian tube data was obtained from a public domain source (Tone et al., 2008) which was obtained using a different platform and had to be integrated with the in-house ovarian cancer initiating cell data. This process included steps for the integration and removal of the platform bias induced due to the use of different platforms between the two datasets. This process has been explained in detail in the later sections. The overall approach for testing the second hypothesis can be seen in the Figure-2.

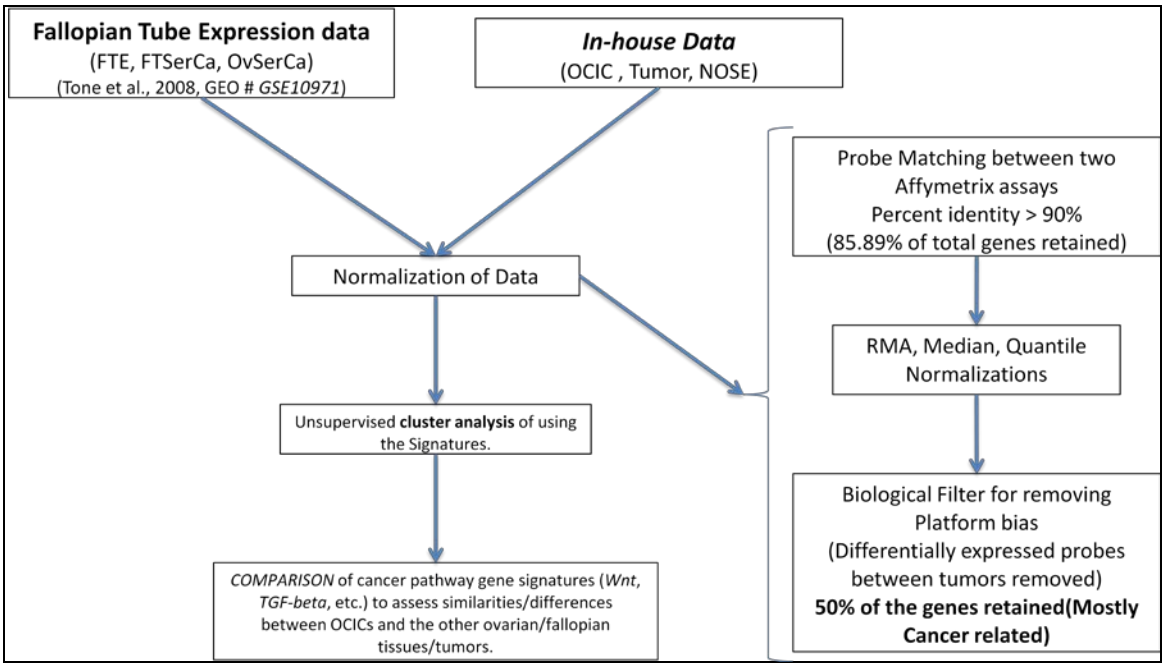

Figure 3: Comparison of Ovarian cancer initiating cells with Fallopian Tube cells.

After the integration of the two datasets, the comparison was performed using the gene signatures identified in the first hypothesis to identify the similarities and differences between the tumors compared to the fallopian epithelial cells and the ovarian surface epithelial cells and also to observe the classification of OCICs with respect to these phenotypes.

Data Collection

The identification and collection of high quality datasets was of utmost importance for correct evaluation of the proposed hypotheses. The samples used in this study were taken mainly from three sources. One of these sets belonged to the samples prepared at Indiana University from the Nephew lab consisting of Ovarian cancer initiating cells and ovarian surface epithelium, here indicated as in-house dataset; the other belonged to a study by Tone et. al which studied the Fallopian Tube epithelium and Fallopian Tumors. The third data set was obtained from the data made available by The

Cancer Genome Atlas consortium for ovarian tumors. The details of these data sets will be explained here.

In-house Dataset

The tumor samples were obtained from patients with stage III serous adenocarcinomas. These samples were minced, suspended in DMEM/F12 medium (Invitrogen), and mixed with 300 units/mL of both collagenase (Invitrogen) and hyaluronidase (Calbiochem), which was followed by overnight incubation (37^{o}C, 5% CO_2). These enzymatically disaggregated suspensions were then filtered (40-µm cell stainer) and washed twice with PBS, and RBCs were removed by Histopaque-1077 (Sigma). As a result, single tumor cells were obtained which were placed under stem cell conditions by re-suspension in serum-free DMEM/F12 supplemented with 5 µg/mL insulin (Sigma), 20 ng/mL human recombinant epidermal growth factor (EGF; Invitrogen), 10 ng/mL basic fibroblast growth factor (bFGF; Invitrogen), and 0.4% bovine serum albumin (BSA; Sigma), which was followed by culturing in Ultra Low Attachment plates (Corning) and subsequent organization into spheres (Zhang, et al., 2008).

The ovarian cancers initiating cells were obtained by fluorescence-activated cell sorting tumors. This was carried out by disassociating small pieces of tumors into single cells followed by washing and removing RBCs. Cells were then suspended in 2% BSA/PBS and labeled with anti-CD44, anti-CD117, and secondary antibodies which were phycoerythrin-labeled and FITC-labeled respectively. The isolation of $CD44^{+}$, $CD117^{+}$, or $CD44^{+}CD117^{+}$ cells was performed using FACSAria flow cytometer and analysed by WinMDI. The cells were sorted twice to assess the OCIC enrichment and purity. Further, the normal ovarian surface epithelial cells were obtained from consented

individuals who showed no signs of ovarian cancer. These methods were all performed at Indiana University in the Nephew lab and are also explained in a previous study by Zhang et.al (Zhang, et al., 2008).

The samples belonging to the three phenotypes explained above i.e. Tumor, Ovarian Cancer Initiating Cells and normal Ovarian Surface Epithelium were further used for generating high-throughput Gene Expression and Methylation data. RNA was obtained from these samples using the standard Qiagen RNeasy purification kits. The RNA extracted from three samples each for Tumor and OCIC (tumor and OCIC from each patient i.e. 3 patients) where as five samples for normal OSE were selected for generating the gene expression data using the Affymetrix gene-1.0-st arrays. The standard procedures for the quality and quantity assessments from Affymetrix were followed. The intensity data obtained from these arrays was further processed for dye-bias and normalization techniques explained in further sections. Further tumor and OCIC samples from the same three patients used for gene expression studies were used for obtaining DNA methylation profiles. Methylation data was also obtained for two normal Ovarian Surface Epithelium samples. The genomic DNA was obtained using standard Qiagen DNeasy purification kits and was subject to sodium bisulfide conversion, labeled with fluorescent dyes and then hybridized to Illumina HumalMethylation 27K bead arrays. The methods and reagents for the sodium bisulfide conversion and array hybridization were performed based on the standard Illumina protocols. The methylation intensity data obtained was further processed for dye bias and normalization techniques explained in further sections.

Public Domain Data

Fallopian Tube Datasets

The data was obtained from a study conducted by Tone et. al (Tone, et al., 2008). The study collected samples for normal fallopian tube epithelial cells (FTEn) from individuals who did not show adnexal malignancy or family history and also did not test positive for BRCA1/2 mutations. They further collected samples for thirteen serous carcinoma samples (SerCa) out of these, six samples showed tubal (FTSerCa) and seven showed ovarian (OVSerCa) origin. RNA from the study cases was obtained using NuGEN Biotin Ovation kit (NuGEN Technologies). Further the RNA was subject to cDNA synthesis which was amplified and purified, enzymatically fragmented and labeled with biotin. The quality and quantity was assessed based on the Affymetrix standard protocols and then the purified labeled cDNA product was hybridized to Affymetrix GeneChip U133A Plus 2.0 arrays. The raw data intensities were deposited in National Center for Biotechnology Information Gene Expression Omnibus resource under the accession number GSE10971. This data was obtained from GEO for the further analysis in this thesis. The obtained raw data was integrated with the in-house data followed by multiple normalization steps explained in later sections.

The Cancer Genome Atlas Dataset

The samples in TCGA are collected from newly diagnosed patients with ovarian serous adenocarcinoma who were undergoing surgical resection and had received no prior treatment for their disease including chemotherapy or radiotherapy. DNA and RNA were isolated from the tissue using an AllPrep DNA/RNA mini kit (Qiagen). The datasets used in this thesis are the raw data from TCGA representing a total of 584 samples that were used for obtaining gene expression genome wide expression data whereas 534

samples were used for obtaining DNA methylation profiles. The gene expression profiles were obtained by using Affymetrix HT Human Genome U133 Arrays. The DNA methylation profiles were obtained by using Illumina HumanMethylation 27k Arrays similar to the platform used for the in-house dataset. These datasets were further subject to integration with the in-house datasets followed by multiple normalization procedures in order to assess the quality of the in-house dataset. This will be explained in the later sections.

Quality Analyses

Gene Expression

As explained earlier, the in-house dataset used the Affymetrix gene-1.0-ST arrays for obtaining the gene expression profiles for the tumor, OCIC and normal OSE phenotypes. A number of normalization approaches were considered for the normalization of the raw data. MAS5 algorithm from Affymetrix adjusts the perfect match (PM) probes by effectively subtracting the signal found on the mismatch (MM) probes. This subtraction is done carefully to avoid producing negative signals and to minimize the effect of outliers. However in the case of Affymetrix gene-1.0-st arrays, it does not report the information for mismatch (MM) probes, instead all the probes in this array are considered perfect match (PM) which gives it the ability of genome wide coverage. Thus, Robust Multichip Average (RMA) approach is suitable as it does not involve an implicit subtraction of the MM probe values. Instead, RMA looks at the distribution of the PM probe values and fits a combination of two distributions, a "noise" distribution that is normally distributed, and a "signal" distribution that is distributed like an exponential distribution. The normalized values are estimated through the expected value of the signal distribution. The raw data was thus normalized using the default RMA

algorithm which involves RMA background correction, Quantile normalization and summarization of the intensities by the Median Polish method were performed. The summarized signals were then log base 2 transformed for each probe set.

The normalized datasets for in-house samples were further evaluated using principal component analysis (PCA) using Partek. The PCA indicated that the samples representing the same phenotype were well correlated with each other and at the same time, the three phenotypes were well divergent from each other. This can be seen in Figure-3.

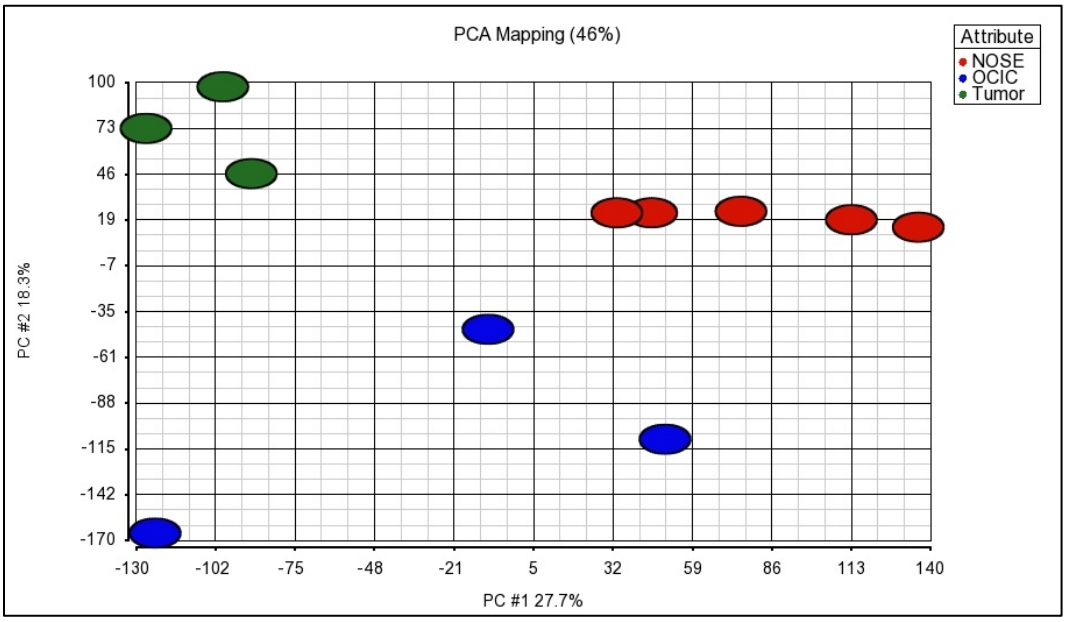

Figure 4: Principal Component analysis of gene expression profiles for the in-house dataset.

The integration of the fallopian tube dataset that used the Affymetrix Human Genome U133 Plus 2.0 Arrays however involved many subsequent steps. After the RMA normalization and the steps described above, the integration step involved the mapping of probes between the two platforms. This was achieved using the sequence matching

information retrieved from Affymetrix depicting the similarity of probes between the two platforms. Only that probe match information was used which was classified as good match by Affymertix (The sequence matching between the two probes showed percent identity greater than 90%). The integration retained 85.89% of the total genes present in both the platforms. The distribution of data after the integration of the two platforms was plotted and can be seen in the Figure-4.

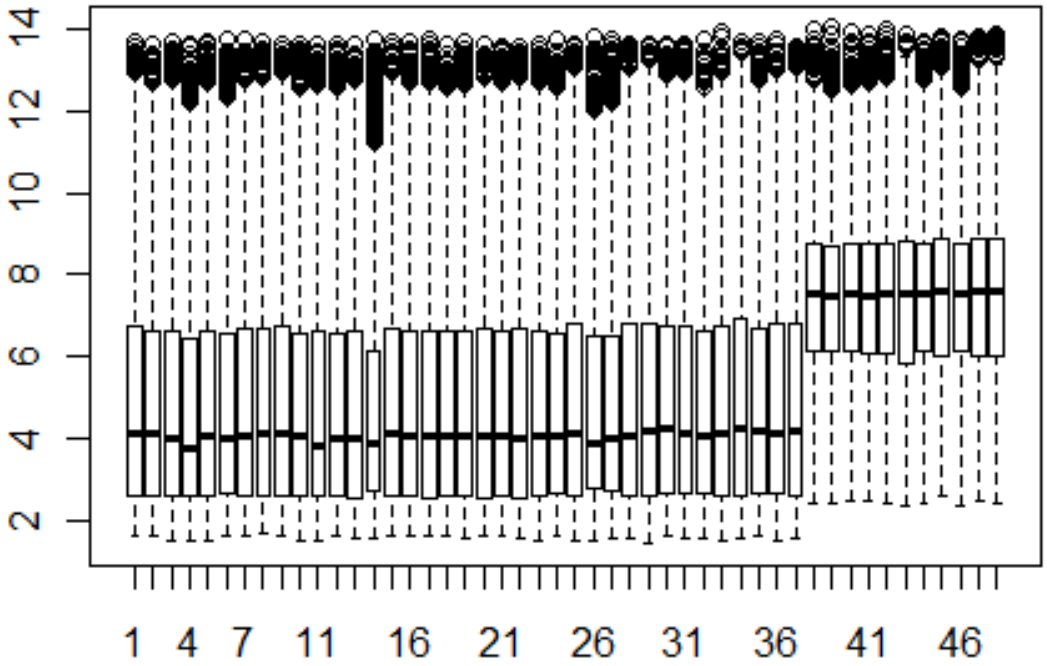

Figure 5: Plot showing the distribution of the data after the integration of the Fallopian Tube data (1 to 37) and in-house (38 to 48).

A clear indication of the presence of a platform bias could be seen from the integrated data. In order to normalize for this platform bias we employed series of statistical and biological validation filters. In order to normalize the datasets statistically we first employed Median normalization in order to center the intensities from the two arrays on a common median. Further assuming that the two arrays have common

distribution of intensities, we employed quantile normalization. The distribution achieved after each of these normalizations can be seen in Figure-5 and Figure-6 respectively.

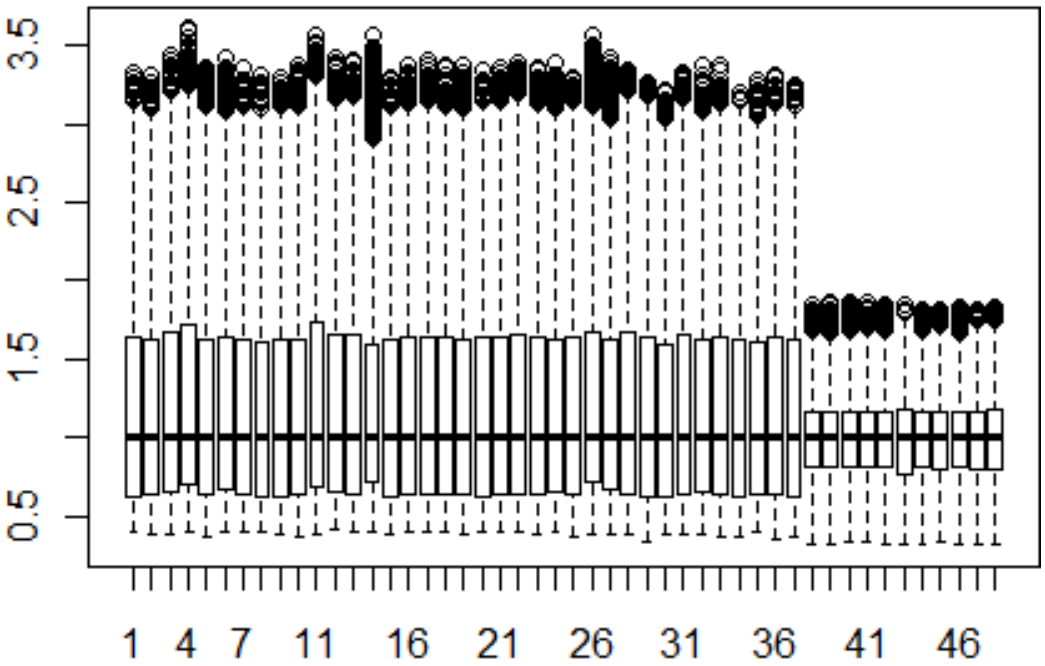

Figure 6: Plot showing the distribution of the integrated data after median normalization showing Fallopian Tube data (1 to 37) and in-house (38 to 48).

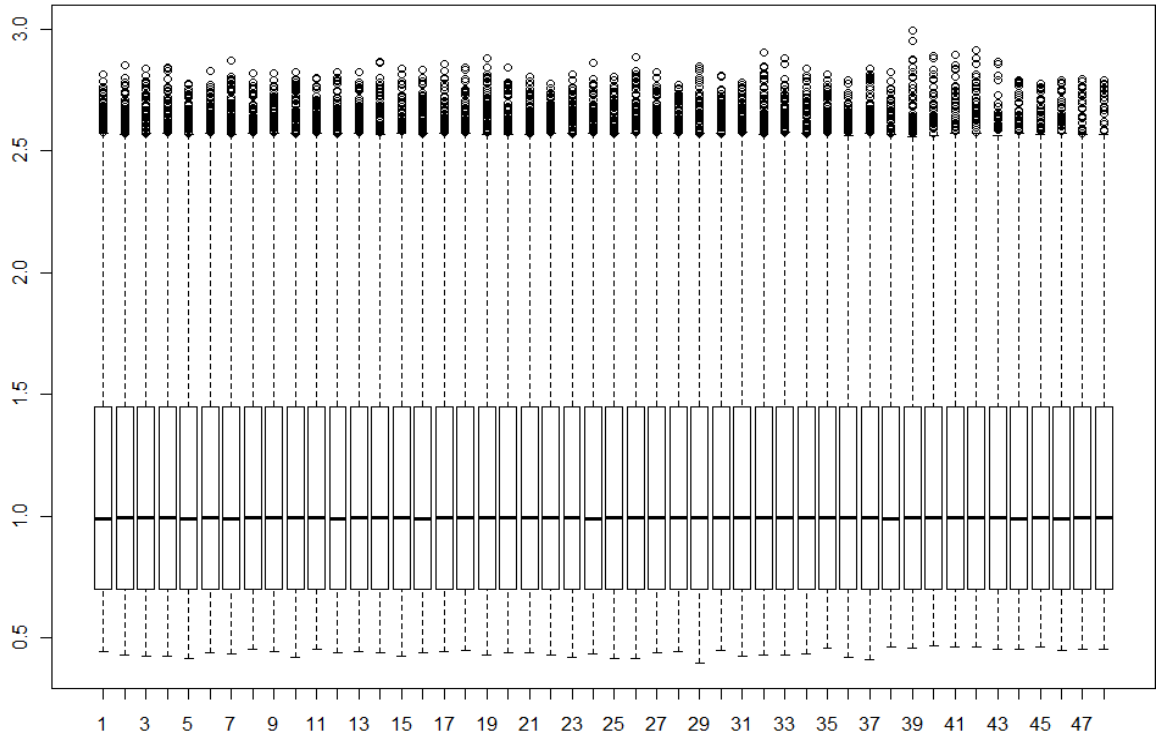

Figure 7: Plot showing the distribution of the integrated data after Quantile normalization showing Fallopian Tube data (1 to 37) and in-house (38 to 48).

The data obtained after quantile normalization seemed to be evenly distributed across the samples for both the arrays. However, to confirm the removal of platform bias from the further analysis, we increased the stringency by applying a biological filter. In order to do this, we determined the ovarian serous carcinoma (OVSerCa) samples from fallopian tube dataset and the Tumors form in-house dataset belonged to the same phenotype i.e. high-grade serous ovarian carcinoma (HG-SOC). Based on this fact, biologically, there should be no differentially expressed probes between these two samples and the ones which are differentially expressed can be considered as a platform bias and should not be included in further analysis. Taking this into consideration, we performed a differential expression analysis on the quantile normalized data for the OVSerCa and in-house Tumor phenotypes. This was done using a two sided Student's T-

test. The probes having a p-value less than 0.05 were considered significantly differentially expressed and were removed from the further analysis. The remaining probes represented 50% of the total genes originally covered by the two platforms. However, a Gene Ontology analysis on the 50% genes removed from further analysis indicated most of the enriched biological process were involved in development of differentiation processes which can be thought of mostly the properties of housekeeping genes. The remaining 50% of the genes in that were included in the dataset can be said to be the important indicators of the differences in the phenotypes in the progression of cancer.

In order to test the consistency of this approach, we used similar normalization steps described above to integrate the in-house tumors with 584 TCGA tumor samples which also match the biological phenotype being ovarian serous adenocarcinoma. After the integration and subsequent normalization procedures, we performed unsupervised hierarchical clustering (using Multi-Experiment Viewer) separately for the in-house data integrated with fallopian tube dataset and for the in-house data integrated with the 584 TCGA tumors. The clustering patterns obtained for the two comparisons can be seen in the Figure-7 and Figure-8 respectively.

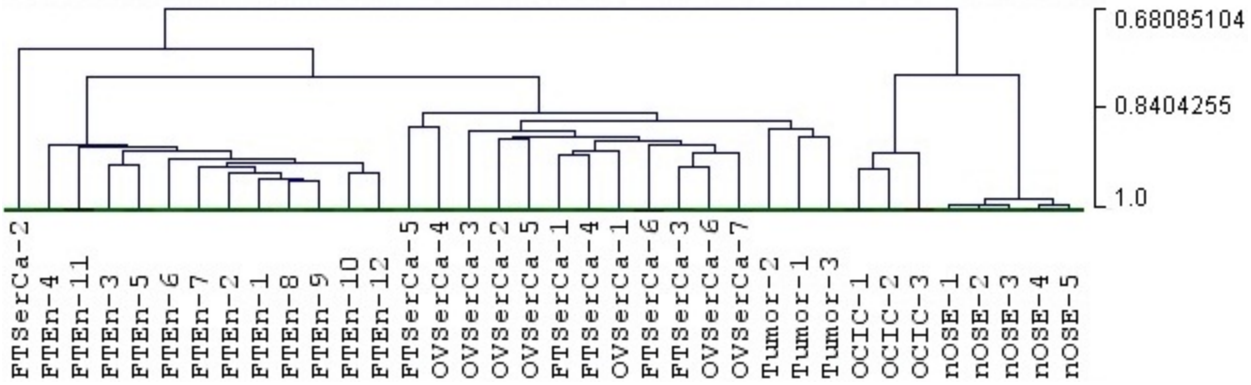

Figure 8: Unsupervised cluster analysis for the in-house data and fallopian tube dataset.

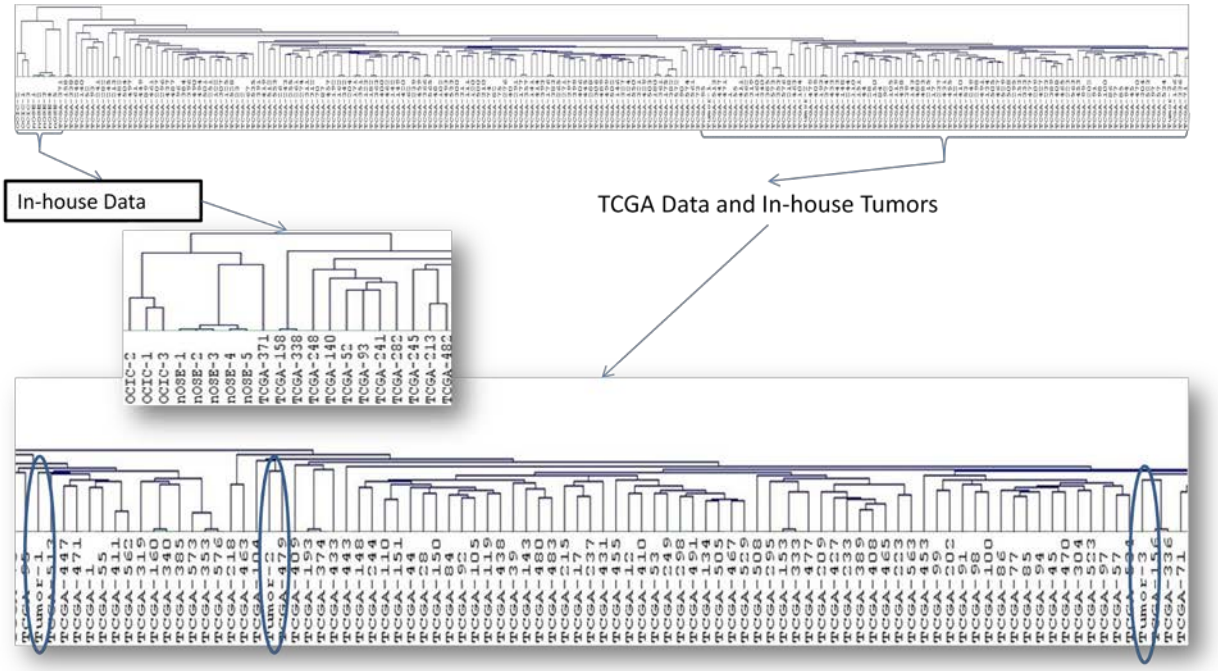

Figure 9: Unsupervised cluster analysis for the in-house data and 584 TCGA Tumors.

From the above clustering patterns observed for the two separate analyses, we can see that as the tumors consistently cluster together with the tumors from the public domain datasets, OCICs are seen to cluster consistently with normal OSE cells. This indicated that even though this stringent normalization approach removes significant

amount of genes from further analysis, the remaining genes will be significant and efficient as indicators for differentiation between phenotypes.

DNA Methylation

The studies of the methylation profiles were performed using the Illumina HumanMethylation 27k platform. This platform provides the intensities in the methylated and unmethylated channels based on the methylation content in the CpG islands. This data was further normalized using the standard normalization algorithms proposed by Illumina using the normalization functions in the Methylumi Package in R programming language (Du, Kibbe, & Lin, 2008). The normalization itself is carried out in a number of steps. The data is first normalized for dye bias, followed by median normalization by inspecting the median intensities in methylated and unmethylated channels at very low and very high beta values. Finally, a new beta value is calculated as a ratio of methylation intensity and the sum of the intensities in the methylated and unmethylated channels. This beta value ranges from 0 to 1, where the values closer to 0 indicate unmethylated CpG islands while the values closer to 1 indicate methylation in the CpG islands. The biological interpretation can be thought of as the value of 0 will indicate both the alleles are unmethylated, a value of 0.5 will indicate that one of the alleles is methylated and value closer to 1 indicates both the alleles are methylated. The distribution of the beta values for the Tumor samples can be seen in the Figure-9.

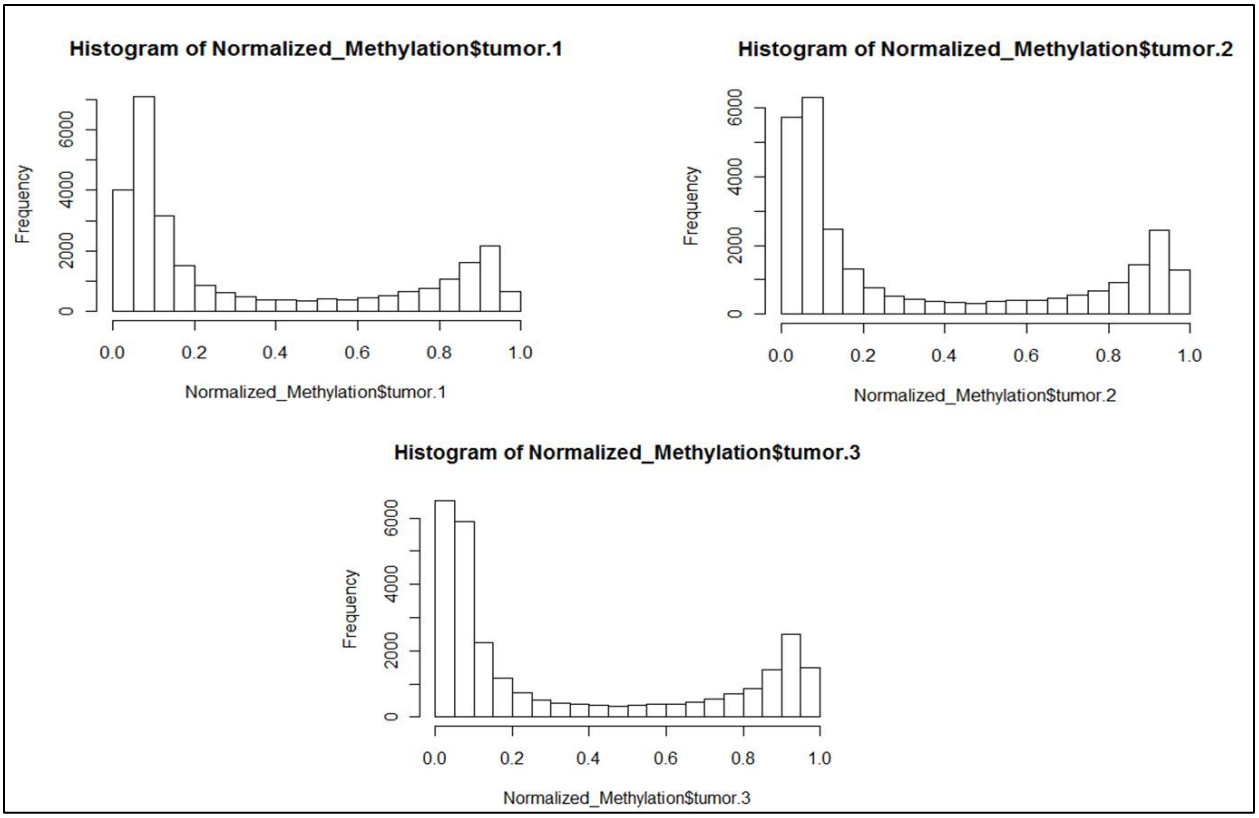

Figure 10: Distribution of Beta values for the Tumor samples.

The normalized datasets for in-house samples were further evaluated using principal component analysis (PCA) using Partek. The PCA indicated that the samples representing the same phenotype were well correlated with each other and at the same time, the three phenotypes were well divergent from each other. This can be seen in Figure-10.

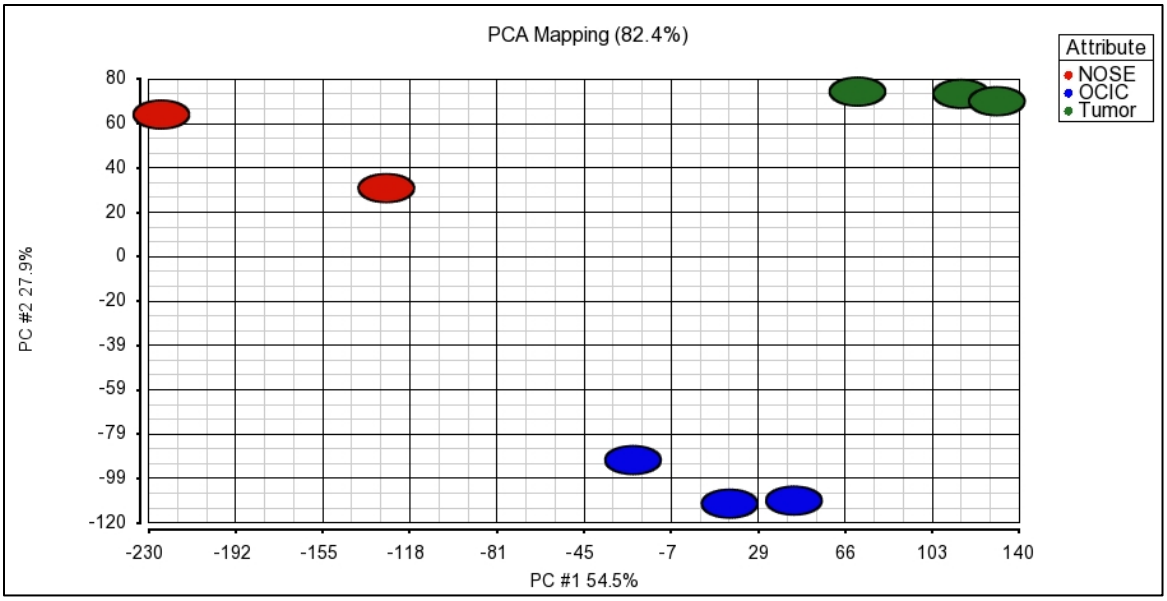

Figure 11: Principal Component Analysis of the Methylation data representing three phenotypes.

In the case of TCGA methylation datasets, studies were carried out using the same platform used here and similar normalization approach was used for these datasets. The TCGA included methylation profiles for 534 ovarian serous carcinoma tumors which were integrated by matching with the probes from the in-house dataset. In order to test whether the profiles for in-house tumors represented the ovarian serous carcinoma phenotype, an unsupervised cluster analysis was performed for this integrated data. This can be seen in the Figure-11.

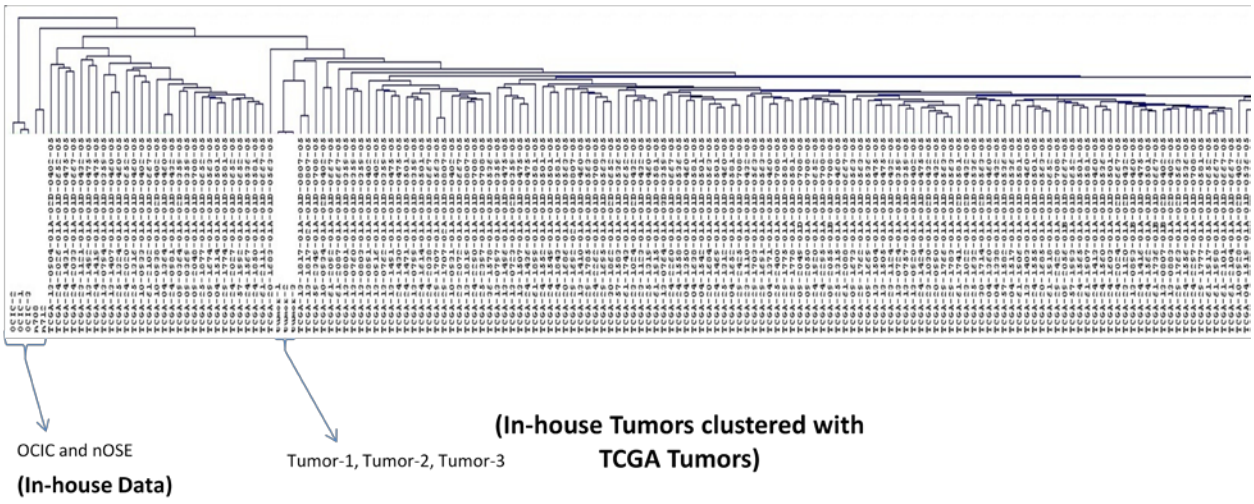

Figure 12: Unsupervised hierarchical clustering of in-house dataset with TCGA dataset.

It was seen that the ovarian tumor methylation profiles for in-house tumors clustered with TCGA tumors indicating that the in-house tumors clearly represent the ovarian serous carcinoma phenotype. Moreover, it was observed that the OCIC and normal OSE showed completely distinct methylation profile compared to the tumors.

Statistical Analyses

Various statistical approaches were used for both the methylation and gene expression data in further analyses. These included differential comparisons between the samples, Gene ontology analysis, Gene set enrichment analysis, unsupervised hierarchical clustering analysis. The details on each of these techniques will be explained here.

Differential Analysis

Differential analysis was carried out for both the DNA methylation and gene expression datasets. However, the stringency criteria differed for each of these datasets based on the nature of the data. The normalized gene expression data was used for the

comparisons of differential expression using two sided Student's T-test over each comparison i.e. OCIC Vs nOSE, OCIC Vs Tumor and Tumor Vs nOSE. For each of the comparison fold change values were calculated separately from the raw data. Genes were then filtered using a T-test p-value where probes with p-value less than 0.01 were selected, then followed by twofold change cutoff, where the previously filtered genes were scanned for the genes which had a fold change difference of greater than 2. A similar analysis was carried out for the methylation data with the inclusion of another filter which checked the methylation status for the pair in the comparison. For this filter, a beta cutoff of 0.5 was used (determined based on the information from the literature and the distribution of beta intensities in the samples), where all the beta values above 0.5 were considered hypermethylated and those below 0.5 were considered unmethylated. For the differential methylation analysis it was confirmed that if one of the pairs of intensities in the comparison is hypermethylated then the other should be unmethylated and vice versa.

Gene Ontology (GO) and Gene set enrichment analysis (GSEA)

The gene ontology analysis was mostly carried out on the datasets that were selected for further functional analysis in the various approaches. On the other hand gene set enrichment analysis was carried out in order to test the enrichment of the gene sets and individual genes in that particular gene set that were selected for further studies. The GO analysis was mainly carried out using default GO settings used in DAVID functional annotation tool (Huang da, Sherman, & Lempicki, 2009). The gene set enrichment analysis was carried out using GSEA tool from the Broad institute (Subramanian et al., 2005).

Unsupervised Hierarchical Clustering

The unsupervised hierarchical clustering has been extensively used in this thesis. All the clustering results were obtained by using MultiExperiment Viewer (MeV) a tool that is part of the TM4 Microarray Software Suite (Saeed et al., 2006). In hierarchical clustering, genes with similar expression patterns are grouped together and are connected by a series of branches, which is called clustering tree (or dendrogram). Experiments with similar expression profiles are also grouped together using the same method. Here we used Pearson correlation coefficient to measure the similarity between gene and sample profiles. A Pearson correlation coefficient indicates the relationship between two ordered sets of gene expression data for several different conditions. It indicates both how the two sets are related and the strength of that relationship. For example, if gene A increases over time and gene B decreases proportionally, their correlation value will be -1.0 because they are perfectly divergent. If the two sets were not perfectly divergent, but still diverged, the correlation would remain negative, but would be greater than -1.0. In contrast, if genes A and B increase proportionally over time, then their correlation will be 1.0. If genes A and B have absolutely no relationship to each other whatsoever, their correlation will be 0. Once we have a table of correlation values between all the genes, clustering is performed with the genes having high correlation values grouped together. By retracing the order in which the genes were progressively joined into clusters and by knowing the correlation value of each step, we can map out which genes are related to each other closely and which genes are related only distantly. This is represented graphically as the tree and the scale represents the Pearson correlation coefficient for the genes.

Chapter 5 RESULTS

Gene Expression

Based on the differential expression analysis of gene expression data explained earlier, the filtered genes were classified into groups which were specific to each of the particular phenotypes. A clustering showing these groups can be seen in the Figure-12.

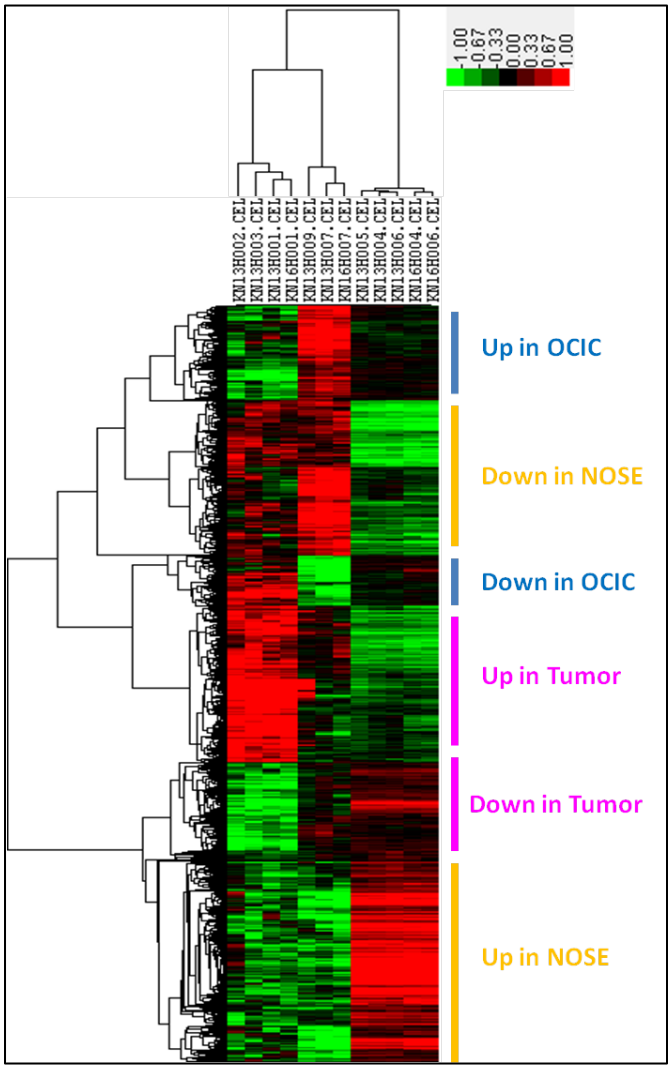

Figure 13: Gene Expression Clustering of the filtered genes.

The genes specific to each of the phenotypes were further studied for the enrichment of specific biological pathways by performing GO analysis. The genes identified in the specific pathways determined from the GO analysis were further used as

gene sets for determining their enrichment on the gene expression data using the gene set enrichment analysis approach on the OCIC compared to rest of the phenotypes.

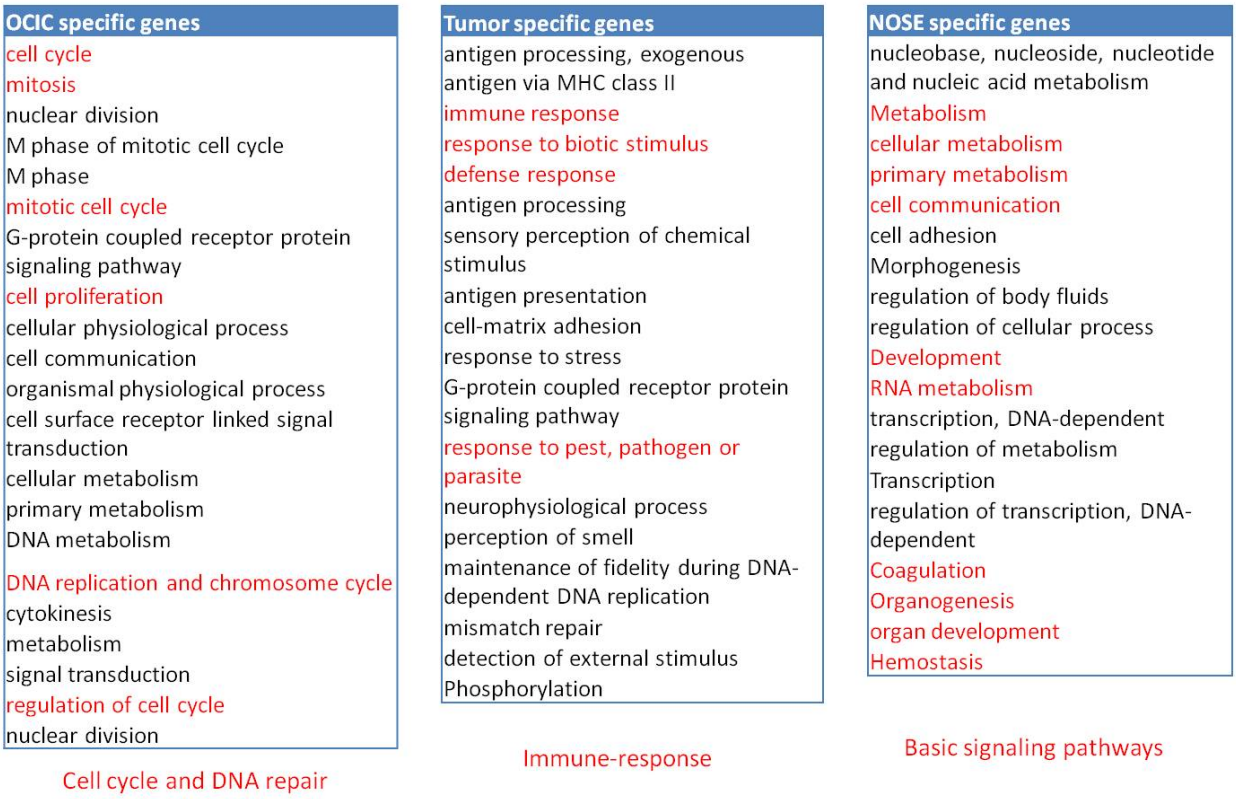

Figure 14: Gene Ontologies for genes specifically up-regulated in each of the phenotype.

The gene ontologies enriched in each of the phenotype were obtained based on the gene specifically up-regulated in each of the tumor, OCIC or normal OSE phenotypes. The ontology terms indicated in red indicate the processes that were found to be unique to that phenotype. These were considered relevant for further analysis approaches. The main terms representing the OCIC specific phenotype were related to Cell cycle and DNA repair processes. For the Tumor specific genes, the enrichment was observed mostly for the terms related to immune-response, whereas in case of normal OSE, most of the processes belonged to the basic signaling pathways. The genes

belonging to the pathways that were considered relevant from the gene ontology analysis were further tested for the enrichment of the gene sets using GSEA.

Enriched in OCIC

NAME	Size	NES
SERUM_FIBROBLAST_CELLCYCLE	123	2.6764326
DOX_RESIST_GASTRIC_UP	40	2.5955234
ZHAN_MM_CD138_PR_VS_REST	37	2.5241997
CMV_IE86_UP	50	2.5168405
LI_FETAL_VS_WT_KIDNEY_DN	148	2.4489558
LEE_TCELLS3_UP	93	2.3930552
YU_CMYC_UP	31	2.382858
CELL_CYCLE_KEGG	85	2.3454473
DNA_REPLICATION_REACTOME	43	2.3266845
MANALO_HYPOXIA_DN	76	2.280617
P21_P53_ANY_DN	44	2.2386074
CROONQUIST_IL6_STARVE_UP	33	2.2316256
CROONQUIST_IL6_RAS_DN	23	2.2220464
VERNELL_PRB_CLSTR1	61	2.2151043
CELL_CYCLE	76	2.20461
IDX_TSA_UP_CLUSTER3	88	2.203463
P21_P53_MIDDLE_DN	23	2.1708589
ADIP_DIFF_CLUSTER5	37	2.1411867
G1_TO_S_CELL_CYCLE_REACTOME	65	2.119298
CANCER_UNDIFFERENTIATED_META_UP	67	2.1040244
ADIP_DIFF_CLUSTER4	34	2.0641186
GOLDRATH_CELLCYCLE	28	2.0599144
HOFFMANN_BIVSBII_BI_TABLE2	188	2.047138
BRENTANI_CELL_CYCLE	77	2.0415292
HSA04110_CELL_CYCLE	112	2.0359983
SHEPARD_GENES_COMMON_BW_CB_MO	60	2.0034811

↑

RED: cell cycle related
BLUE: differentiation related
GREEN: hypoxia related

Enriched in REST

NAME	Size	NES
BAF57_BT549_UP	215	-2.5130732
CARIES_PULP_UP	197	-2.443954
IRITANI_ADPROX_VASC	150	-2.4296489
AGEING_KIDNEY_UP	369	-2.3789291
BOQUEST_CD31PLUS_VS_CD31MINUS_DN	239	-2.301886
CMV_24HRS_DN	70	-2.2725976
ADIP_VS_PREADIP_DN	36	-2.256739
AGEING_KIDNEY_SPECIFIC_UP	170	-2.2435393
HINATA_NFKB_UP	106	-2.239741
TAKEDA_NUP8_HOXA9_3D_UP	162	-2.2237816
CORDERO_KRAS_KD_VS_CONTROL_UP	68	-2.202057
TGFBETA_ALL_UP	79	-2.2005646
CROONQUIST_IL6_STROMA_UP	37	-2.1753893
HTERT_DN	66	-2.1642413
IRITANI_ADPROX_DN	58	-2.1318233
CMV_ALL_DN	101	-2.1297038
WIELAND_HEPATITIS_B_INDUCED	94	-2.1095607
ATRIA_UP	183	-2.10831
TAKEDA_NUP8_HOXA9_10D_UP	161	-2.0899184
DORSEY_DOXYCYCLINE_UP	30	-2.088948
VEGF_MMMEC_12HRS_UP	29	-2.0864842
SANA_TNFA_ENDOTHELIAL_UP	78	-2.0752134
ADIPOGENESIS_HMSC_CLASS8_DN	32	-2.0679107
IFNA_HCMV_6HRS_UP	48	-2.0653877
TAKEDA_NUP8_HOXA9_8D_DN	183	-2.0640814
TGFBETA_EARLY_UP	46	-2.0600245
ADIP_VS_FIBRO_DN	26	-2.0598052
GILDEA_BLADDER_UP	28	-2.05474
HSA04510_FOCAL_ADHESION	191	-2.0499306
VEGF_MMMEC_6HRS_UP	48	-2.0489
BROCKE_IL6	124	-2.0423462
GN_CAMP_GRANULOSA_DN	59	-2.0397344
PASSERINI_GROWTH	33	-2.036107
KRETZSCHMAR_IL6_DIFF	124	-2.0280857
NI2_MOUSE_UP	39	-2.025265
TAKEDA_NUP8_HOXA9_16D_UP	144	-2.0191185
NEMETH_TNF_UP	78	-2.018691

Figure 15: GSEA for the gene sets related to GO enriched processes carried out on entire dataset for the comparison of OCIC with Rest of the phenotypes.

The GSEA revealed the pathways and gene sets identified unique for OCIC were indeed enriched significantly in the OCICs compared to the tumor and normal OSE phenotypes. It was seen that the gene sets related to Cell Cycle, Cell differentiation and Hypoxia pathways were significantly enriched in OCICs.

DNA Methylation

In order to identify the pathways altered specifically in each phenotype especially due to the alteration in the methylation patterns in the three phenotypes, genes that were

hypermethylated specifically in these phenotypes were selected. We considered genes with beta intensities greater than 0.5 to be hypermethylated, thus filtering out the genes that showed unmethylation i.e. beta intensity less than the above define threshold.

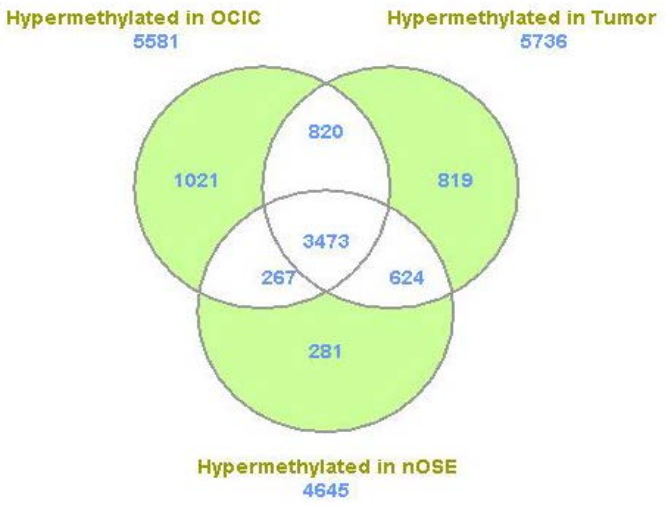

Figure 16: Genes hypermethylated in each phenotype. The genes indicated in shaded regions can be considered specific to that phenotype and were used for further analysis.

The above filtering provided a list of genes for each phenotype indicated in Figure-15. It was observed that a total of 5581 genes were hypermethylated in OCICs out of which 1021 genes belonged specifically to OCICs. In case of Tumors, 5,736 genes were found to be hypermethylated out of which 819 genes were specifically methylated in tumors. Whereas, out of 4,645 genes that were hypermethylated in normal OSE, only 281 were seen to be specifically hypermethylated in this phenotype. These specifically hypermethylated genes in each phenotype were subjected to further gene ontology analysis to identify altered pathways.

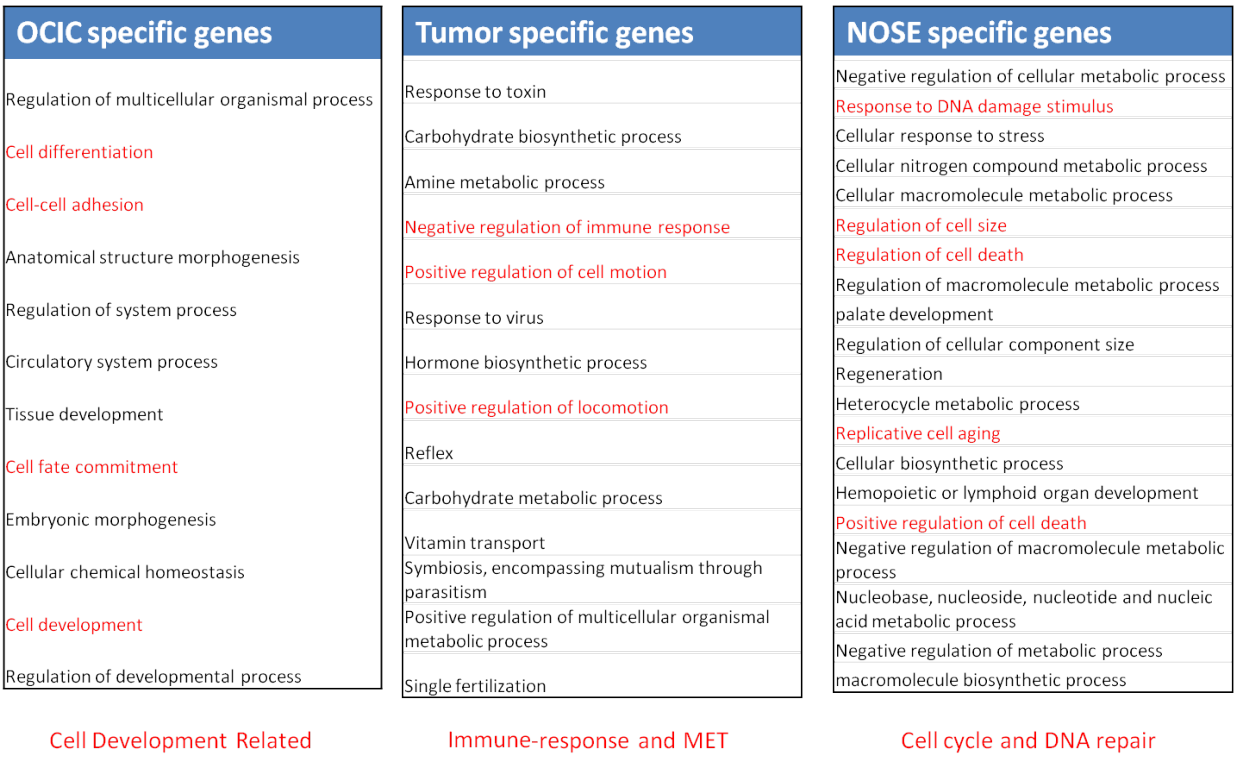

Figure 17: Gene Ontology analysis for specifically hypermethylated genes in each phenotype.

From the gene ontology analysis it was found that the most of the cell development related pathways were hypermethylated in OCICs which can correspond to silencing of their genes and functions. In the case of tumor specific genes, it was seen that most of the immune response and mesenchymal to epithelial transition related pathways were hypermethylated. Also in the case of genes specifically hypermethylated in normal OSE, most of the processes related to cell cycle and DNA repair pathways were seen to be hypermethylated and hence can be thought of as silenced.

Determination of Signature Pathways

Based on the biological processes that were enriched in the expression and methylation analyses explained above, we determined the processes that contributed to various signature pathways that were known to be related to or contributing in the

development of tumor phenotypes in various cancers. The genes belonging to these signature pathways were instrumental in differentiating between the different tumor phenotypes and the cell types known to promote cancer phenotypes. We determined four important signatures from literature that were known to be related to ovarian cancers that are known to show enrichment of the identified gene ontologies and involved in cancer progression.

- Epithelial to mesenchymal transition.
- TGF-beta signaling pathway.
- WNT signaling pathway.
- Notch signaling pathway.

The details about these pathways and their results for the expression and methylation datasets are explained further.

Epithelial – Mesenchymal Transition Signature

As explained earlier, EMT is the process by which an epithelial cell becomes more mesenchymal (Valcourt, Kowanetz, Niimi, Heldin, & Moustakas, 2005). The signature was obtained from a recent study conducted by Taube et al. on human breast cancer cells (Human mammary epithelial cells or HMLE) (Taube et al., 2010). The EMT core signature was obtained by inducing HMLE cells to express genes like GSC, Snail, Twist and TGF-β1 or the knock down of the expression of gene E-cadherin (These genes have been known to be responsible to induce an EMT process in cells eventually converting the epithelial cells to mesenchymal phenotype.). The resulting set of gene expression changes between the two phenotypes was defined as the EMT core signature. This signature included 159 down-regulated and 87 up-regulated genes (showing atleast

2-fold change) which was obtained from all of the EMT inducing signals discussed above. The study also indicated that the use of the total mRNA isolated from entire tumors may preclude the detection of cells that have undergone EMT as only a small proportion of the neoplastic cells in each tumor may exhibit the EMT phenotype. It is also indicated that FOXC1 gene expression correlated with a poor survival among the breast cancer patients.

Unsupervised Cluster analysis

An unsupervised cluster analysis was performed for each of the signature datasets of EMT i.e. the clustering was performed on up-regulated and down-regulated genes separately. The clustering was performed for both gene expression and DNA methylation data. It showed a clear distinction between the normal OSE and tumor phenotypes. The normal OSE showed a clear adherence to the mesenchymal phenotype such that most of the genes that were included in the up-regulated signature were also up-regulated in normal OSE and vice versa (for down-regulated signature). On the other hand, tumor showed a clear epithelial phenotype as most of the up-regulated signature genes were down-regulated in tumor and vice versa (for down-regulated signatures). In case of OCIC, they clustered with normal OSE for both the signature gene sets suggesting an inclination toward mesenchymal phenotype. The clustering patterns for both the signatures can be seen in Figures 18 and 19. In the case of methylation intensities, the clustering patterns for both the signature gene sets showed similarities in the tumor and normal patterns but OCICs seemed to have a great difference in the methylation patterns as more genes were methylated in OCICs compared to tumors and normal cells.

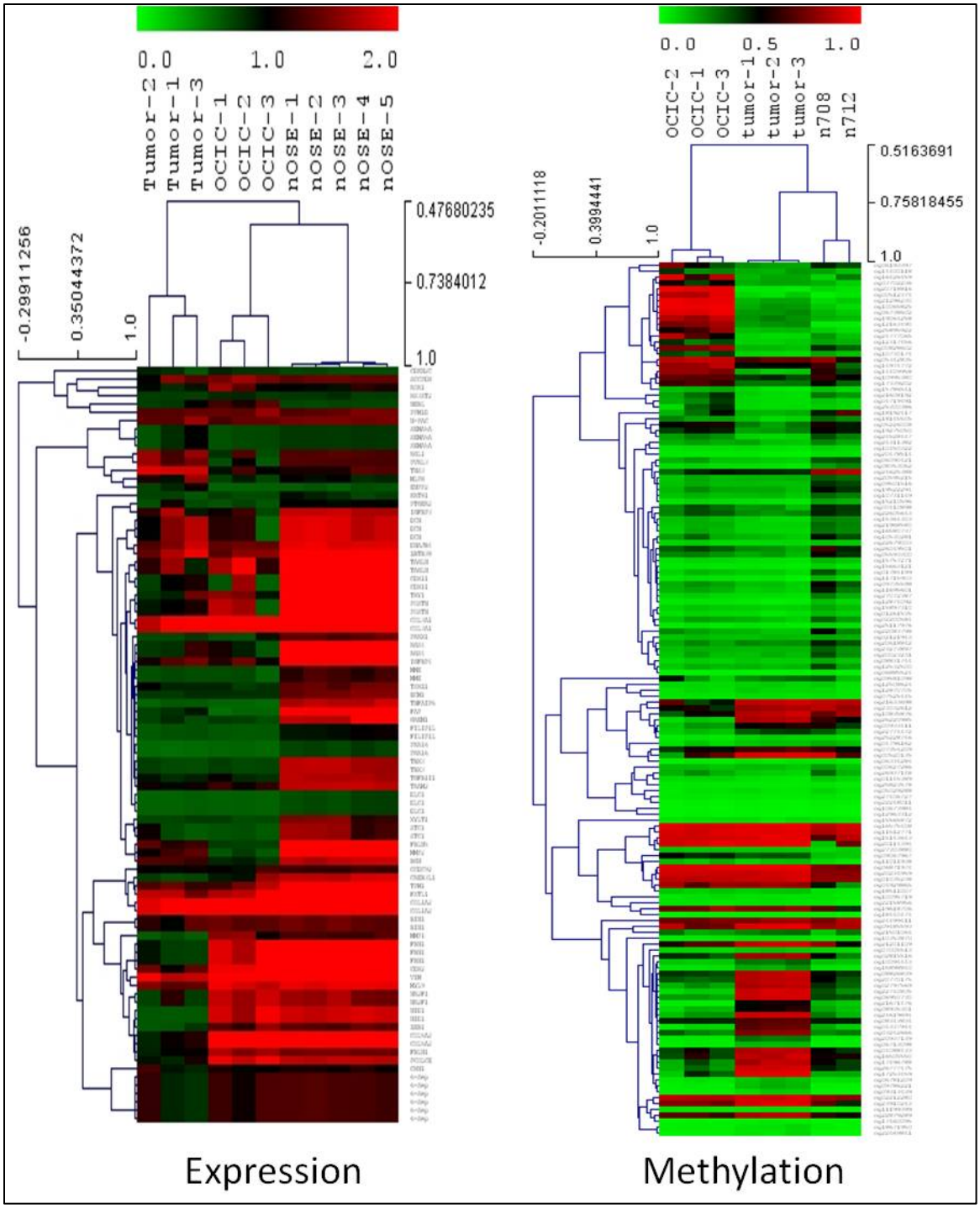

Figure 18: Unsupervised Clustering of Expression and Methylation intensities for genes Up-regulated in EMT signature.

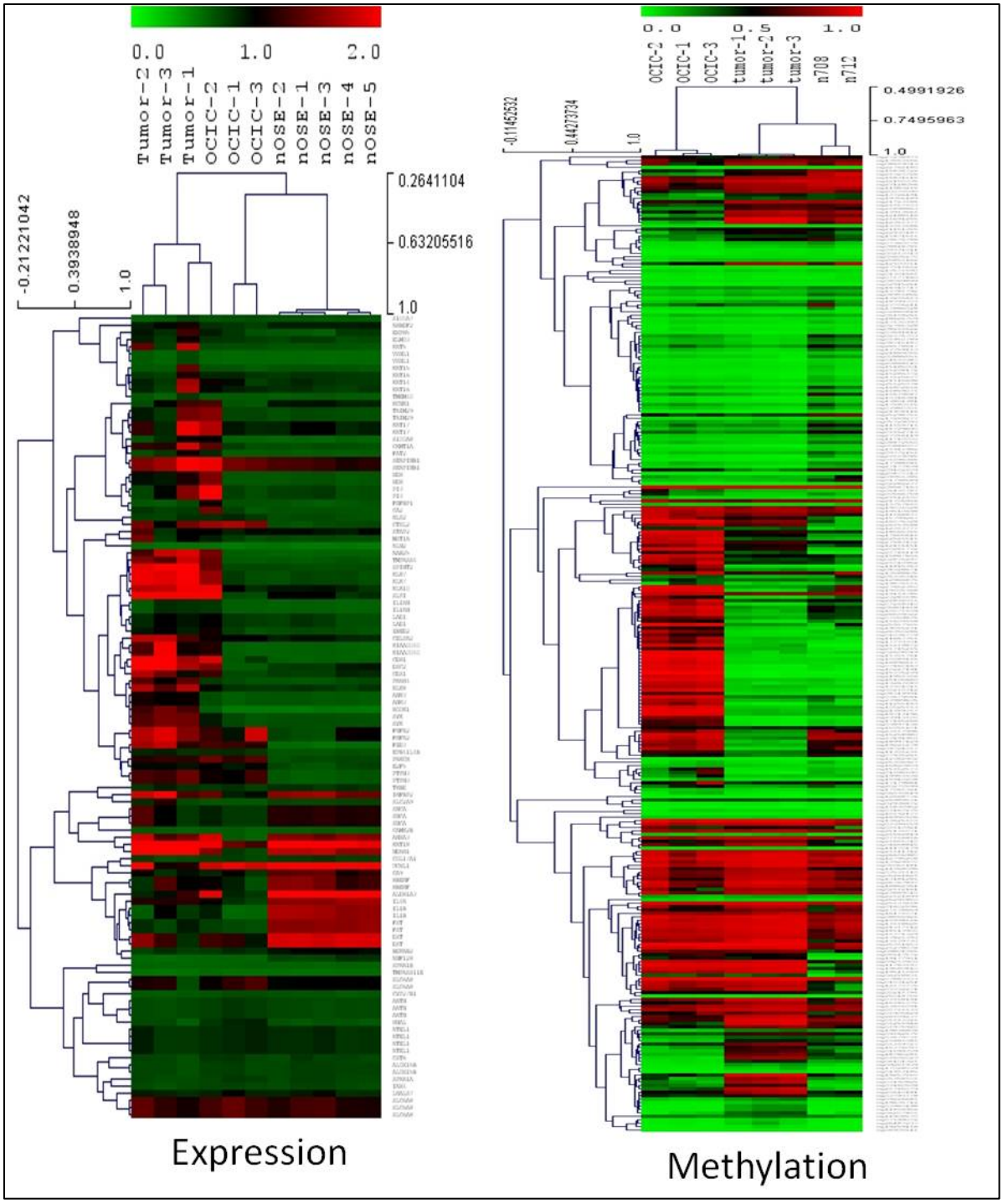

Figure 19: Unsupervised Clustering of Expression and Methylation intensities for genes Down-regulated in EMT signature.

TGF-β Signature

As explained TGF- β is known to be a precursor pathway for the induction of EMT. The TGF- β signature studied here was obtained from a recent study by Valcourt et al. (Valcourt, et al., 2005). This study was carried out on mouse mammary epithelial cells which were stimulated for a considerable amount of time with TGF- β1, to induce the EMT programs, consequently inducing a mesenchymal phenotype. The treated cells and the normal epithelial cells were compared for the gene expression patterns and the genes differentially expressed between the two phenotypes were selected. It was seen that 344 independent genes were regulated by TGF- β1 out of which, 205 (60%) were shown to be up-regulated and 139 (40%) were down-regulated in the EMT induced cells compared to normal epithelial cells. These gene sets were used for the further analysis.

Unsupervised Cluster Analysis

An unsupervised cluster analysis involving the gene expression and methylation data was carried out for both the up-regulated and down-regulated gene signatures. The clustering showed that the tumors and OCICs showed similar clustering patterns however, the correlation of the similarities were low in both the clustering of up-regulated and down-regulated genes. In both the cases, the normal OSE cells showed agreement with the mesenchymal phenotype as the genes up-regulated in the signature were also up-regulated in normal OSE and those that were down-regulated were also down-regulated in normal. The methylation profiles for the clustering analysis in both the signature gene sets showed methylation patterns fairly similar for tumor and normal OSE, whereas the patterns observed for OCICs greatly diverged from the tumor and normal with the number of methylated genes being more in OCICs.

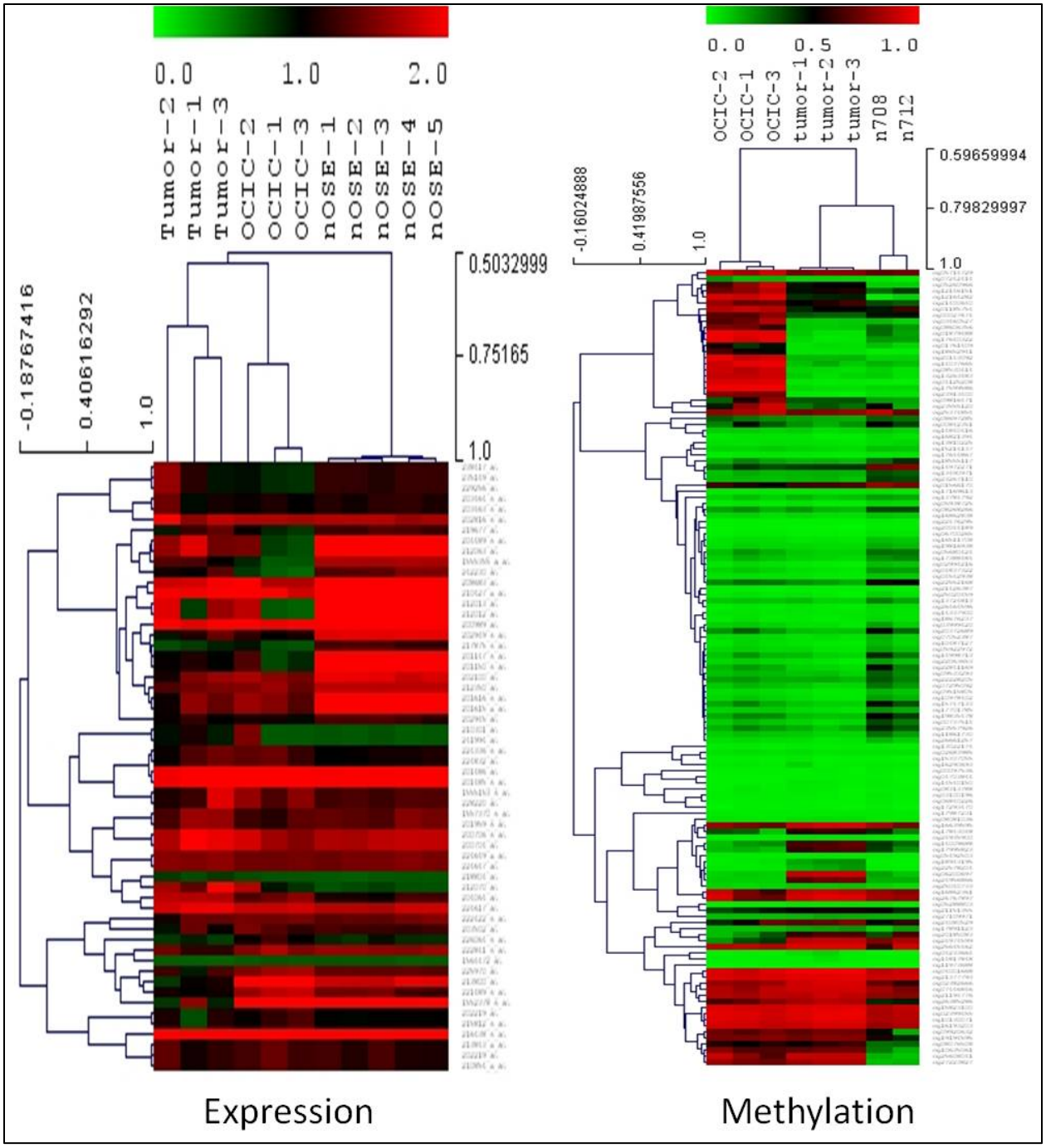

Figure 20: Unsupervised Clustering of Expression and Methylation intensities for genes Up-regulated in TGF-beta signature

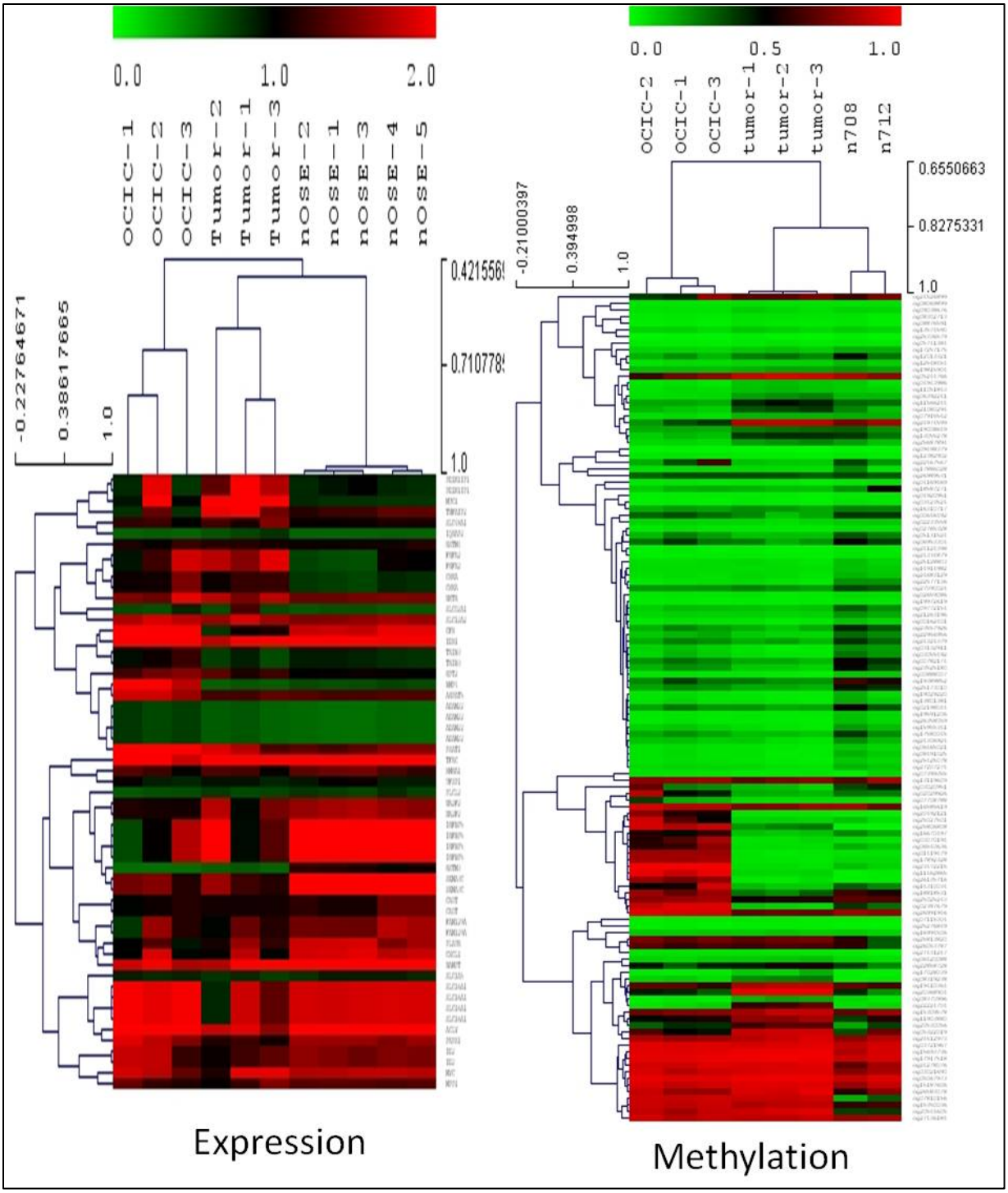

Figure 21: Unsupervised Clustering of Expression and Methylation intensities for genes Down-regulated in TGF-beta signature

The expression patterns of TGF- β were further studied by looking at the genes differentially expressed in the tumor and OCIC phenotypes compared to normal OSE. The genes were selected based on the differential expression criteria (explained in methods section) and also the fold change. This analysis was performed on the genes represented in the TGF-beta signaling pathway annotated in the KEGG database. The details of this analysis can be seen in Figures 22 and 23. It was seen in case of tumors that most of the genes in the upstream of the signaling pathway were down-regulated whereas most of the genes contributing in Apoptosis and Cell Cycle were showed to be up-regulated. A similar expression pattern was seen for OCICs where some of the genes in the upstream of the signaling pathway were up-regulated. The major difference between the tumors and OCICs was that TGF- β was up-regulated in OCIC where as it was down-regulated in tumors.

OCIC vs. NOSE

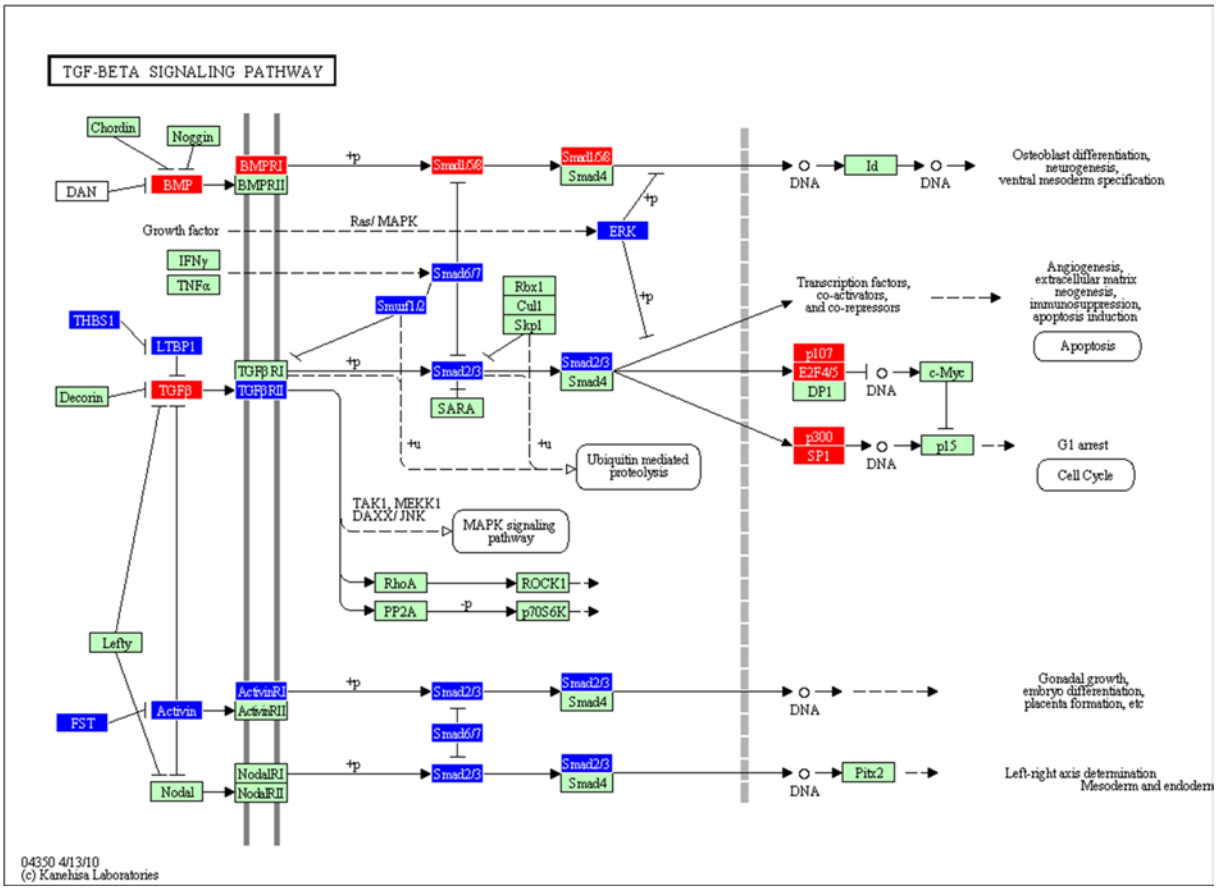

Figure 22: Differential analysis for TGF-beta signaling pathway for OCIC compared to normal OSE.

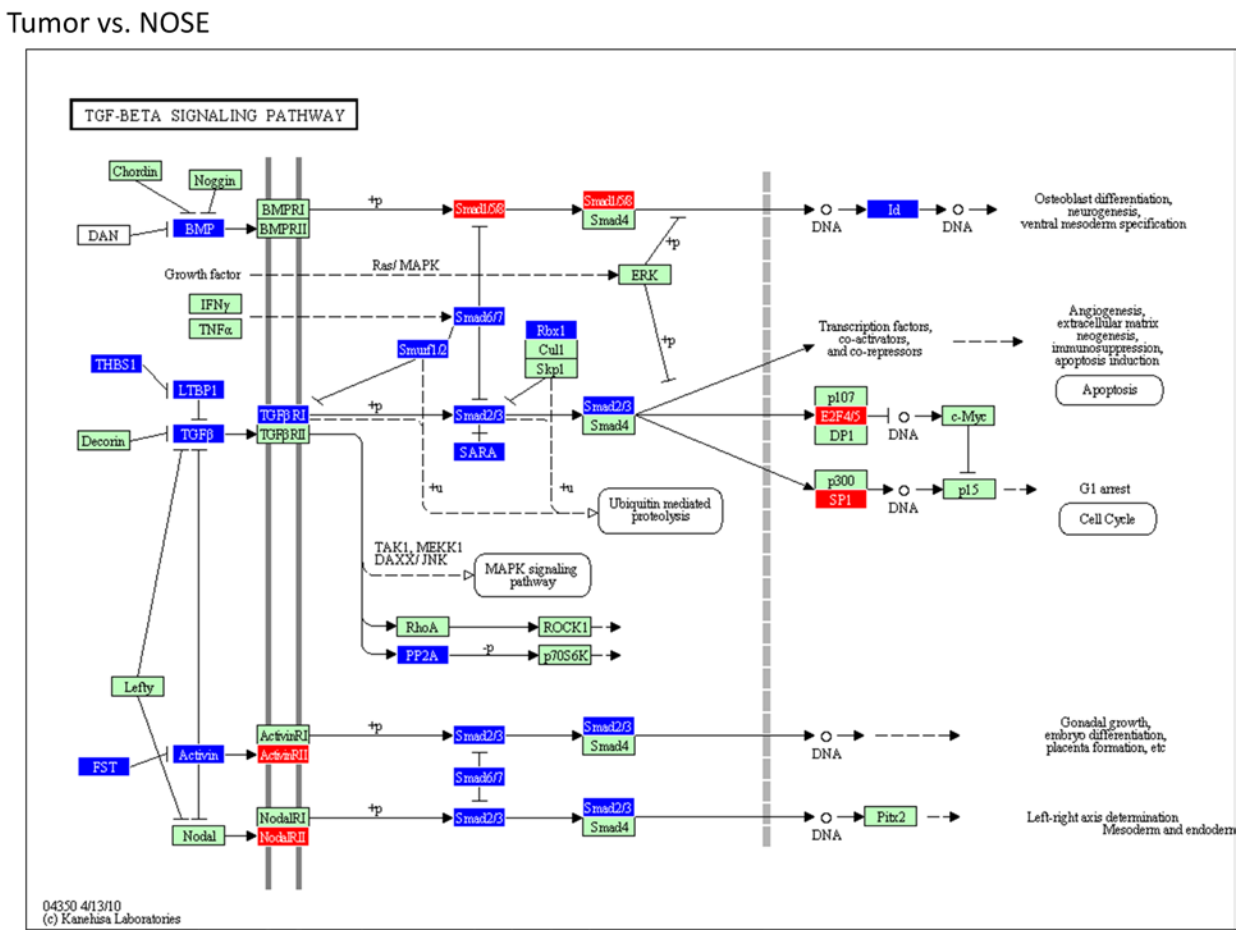

Figure 23: Differential analysis for TGF-beta signaling pathway for Tumor compared to normal OSE.

As TGF-beta regulated wide range of functions like proliferation, apoptosis, differentiation and especially migration, the results obtained in this analysis suggests that the TGF-beta signaling pathway seems to be active in OCIC where as it is inactive in tumor cells. This indicates the OCICs being more predisposed to the properties like migration compared to tumor.

Methylation Signature

Another signature gene set which showed enrichment for biological processes related to TGF-β and EMT signaling was obtained by Matsumura et al. The signature included the

genes proposed in the study to be commonly methylated in cancer. This signature indicated the biological and clinical relevance of DNA methylation in ovarian cancer (Matsumura et al., 2011).

Signature

This signature was obtained form a study of ovarian cancer data. The signature contained 378 genes that were identified by growing cultures in the presence or absence of DNMT inhibitors. It was found that the signature included many TGF-beta related genes and the study suggested that the age related epigenetic modifications lead to suppression of TGF-beta pathway. This was stated based on the hierarchical clustering that was performed on the 378 genes and the corresponding generation of clusters that were strongly correlated with TGF-beta pathway activity shown to be discriminate patients based on age. The cells in the cultures were treated with DNA-hypermethylating agents like 5-aza-2'-deoxycytidine (decitabine or 5-Aza-dc) or 5-azacytidine (5-Azac), followed by gene expression microarray analysis in order to detect genes likely to have been silenced by DNA methylation in cancer. The gene ontologies enriched in this set included adhesion, cell-migration, angiogenesis and immune-response which are all relevant to the functions of TGF-beta super family pathway.

Unsupervised Hierarchical Clustering

An unsupervised hierarchical clustering analysis was performed on the 378 genes found to be commonly methylated in ovarian cancer. From the analysis shown in Figure-24 it was seen that majority of the signature genes were methylated and silenced in OCIC compared to Tumor and normal OSE. Consistent with the clustering patterns observed in previous analysis, methylation patterns for tumors were similar to normal, whereas

OCICs showed a great divergence form the two phenotypes with more number of genes seen to be methylated.

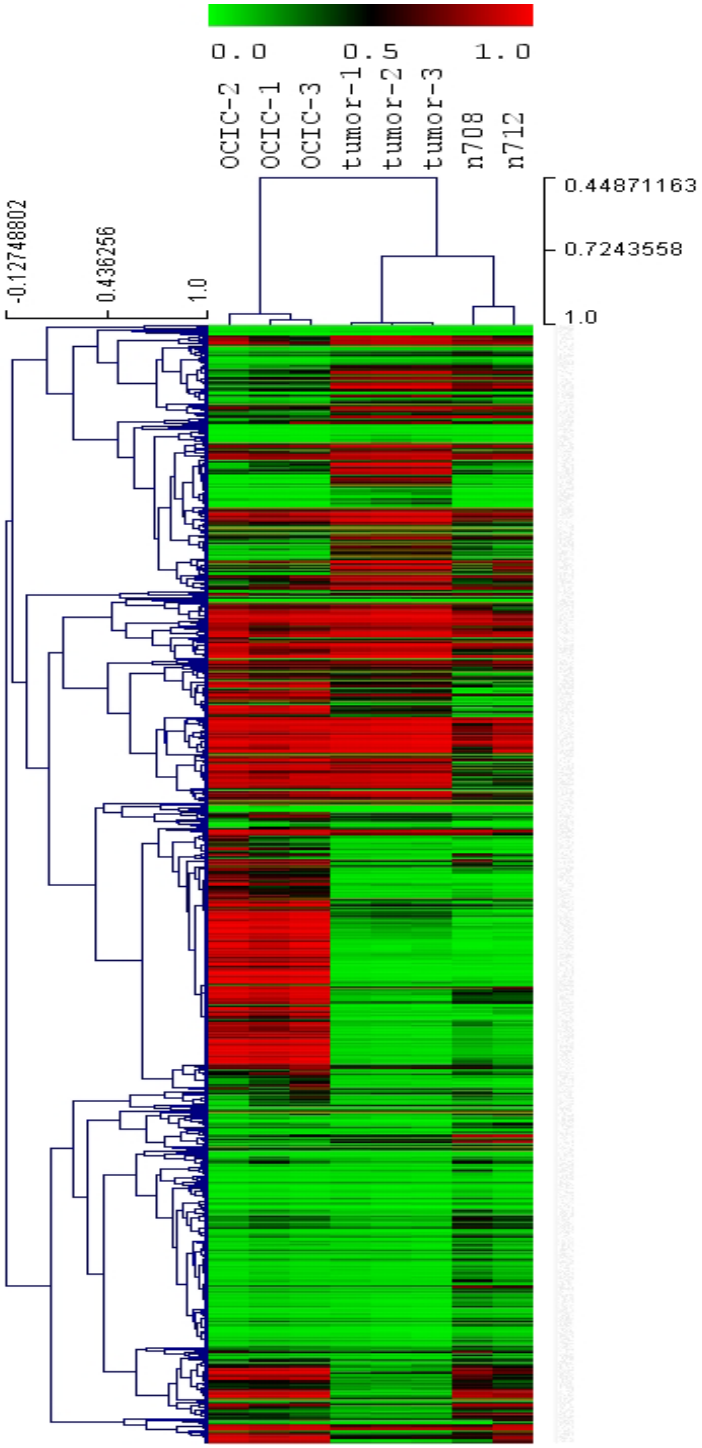

Figure 24: Unsupervised Clustering of genes commonly methylated in cancer.

Wnt signaling pathway

The Wnt proteins are secreted morphogens that are mainly required for basic developmental processes, such as cell-fate specification, progenitor-cell proliferation and the control of asymmetric cell division, in many different species and organs. According to the KEGG annotation, there are at least three different Wnt pathways: the canonical pathway, the planar cell polarity (PCP) pathway and the Wnt/Ca2+ pathway. In the canonical Wnt pathway, the major effect of Wnt ligand binding to its receptor is the stabilization of cytoplasmic beta-catenin through inhibition of the beta-catenin degradation complex. Beta-catenin is then free to enter the nucleus and activate Wnt-regulated genes through its interaction with TCF (T-cell factor) family transcription factors and concomitant recruitment of coactivators. Planar cell polarity (PCP) signaling leads to the activation of the small GTPases RHOA (RAS homologue gene-family member A) and RAC1, which activate the stress kinase JNK (Jun N-terminal kinase) and ROCK (RHO-associated coiled-coil-containing protein kinase 1) and leads to remodelling of the cytoskeleton and changes in cell adhesion and motility. WNT-Ca2+ signalling is mediated through G proteins and phospholipases and leads to transient increases in cytoplasmic free calcium that subsequently activate the kinase PKC (protein kinase C) and CAMKII (calcium calmodulin mediated kinase II) and the phosphatase calcineurin.

Unsupervised Cluster Analysis

The signature gene set for Wnt signaling pathway was retrieved from MSigDB which is a database hosted by Broad Institute (Subramanian, et al., 2005). The gene set included a total of 89 genes which were shown to be integral genes in the Wnt signaling pathway. A hierarchical clustering of both the DNA methylation and gene expression

intensities revealed that most of the genes were similarly expressed in tumors and OCICs. The OCIC samples clustered with tumor samples with a very significant correlation of greater than 0.7 signifies the close similarities in the expression patterns in the two phenotypes. The methylation patterns observed in this signature is consistent with previous analysis suggesting more methylation observed in genes belonging to OCIC compared to tumor and normal OSE cells, whereas the later phenotypes show less differences in the methylation patterns. OCIC on the other hand showed great divergence in the methylation patterns compared to the other two phenotypes.

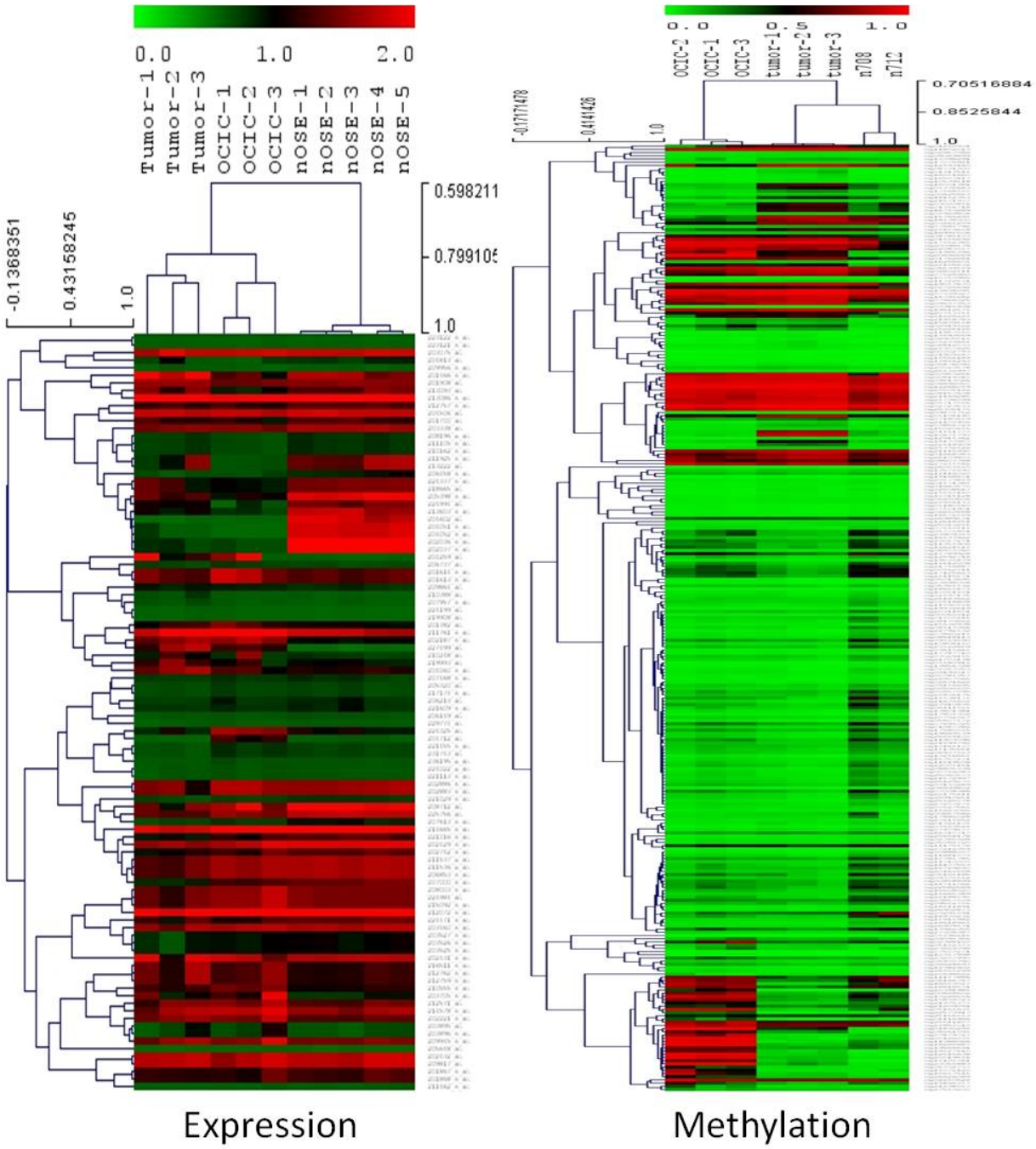

Figure 25: Unsupervised clustering of the genes from WNT signaling pathway from KEGG.

The expression patterns of Wnt signaling pathway were further studied using differential expression in the genes belonging to tumor and OCIC phenotypes compared to normal OSE. The genes were selected based on the differential expression criteria

(explained in methods section) and also the fold change. The details of this analysis can be seen in Figures 26 and 27. Consistent with the unsupervised analysis, the differential expression patterns showed similar gene expression patterns in the Tumor and OCIC cells with some differences in the expression of few downstream genes. Some of these genes which regulate cell cycle process were upregulated in OCIC but down-regulated in Tumors.

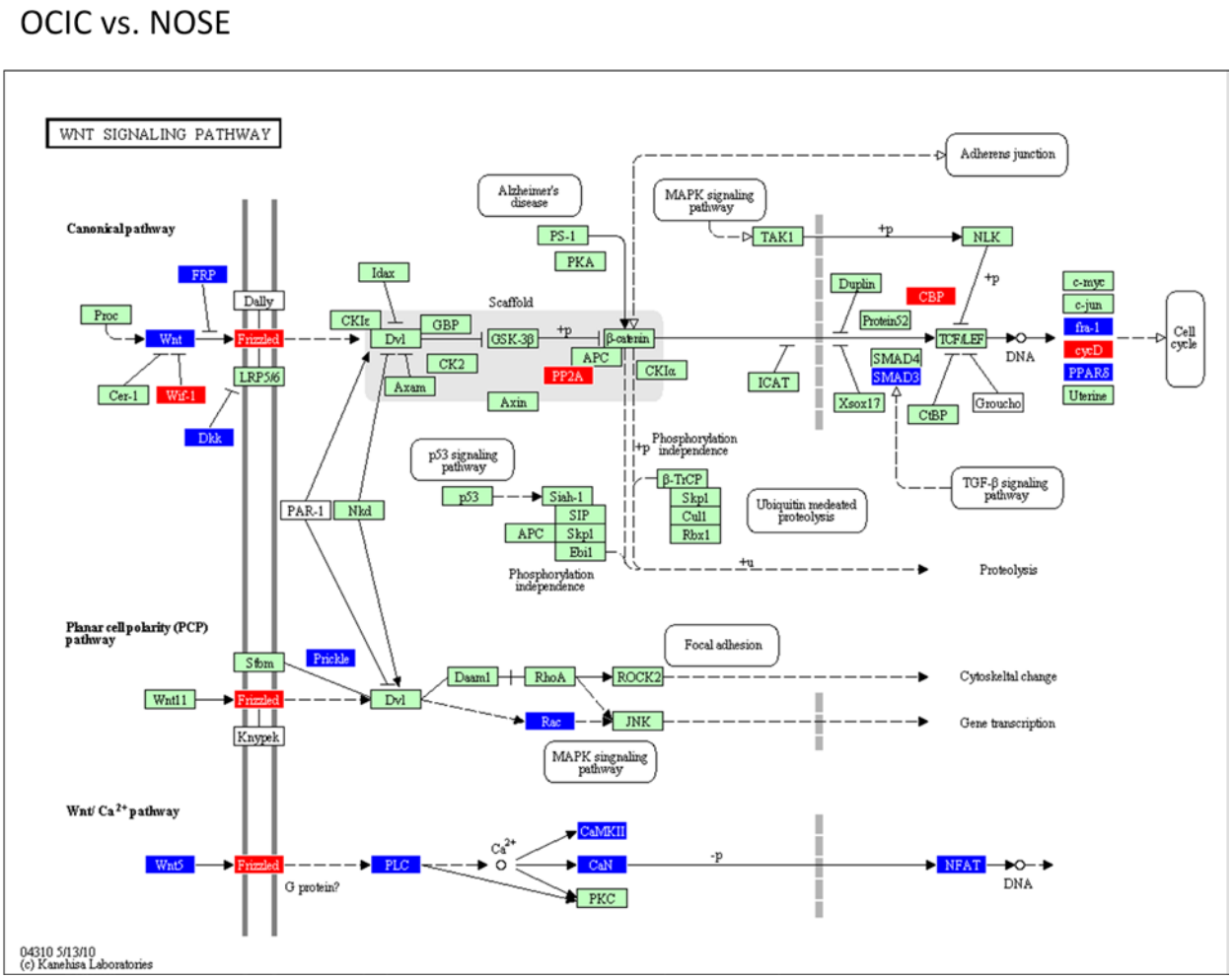

Figure 26: Differential analysis for Wnt signaling pathway for OCIC compared to normal OSE.

Tumor vs. NOSE

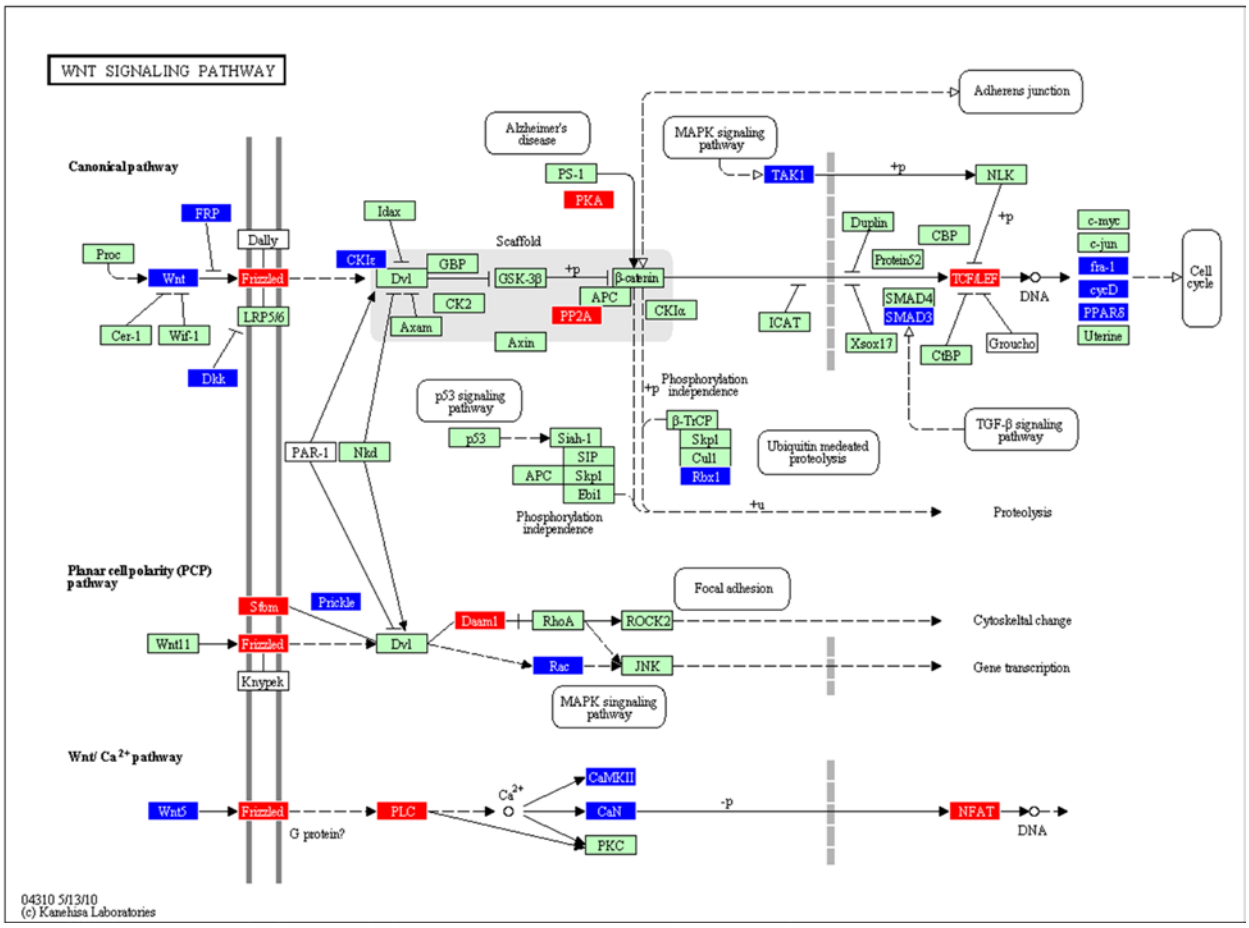

Figure 27: Differential analysis for Wnt signaling pathway for Tumor compared to normal OSE.

Based on the differential analysis it could be said that the Wnt signaling pathway being responsible for many of the basic developmental processes is downregulated in both the tumor and OCIC phenotypes. This suggests that the OCIC share the expression patterns found in the tumor cells with an addition of the upregulation of genes responsible for cell cycle and cell proliferation.

Notch Signaling Pathway

The Notch signaling pathway is an evolutionarily conserved, intercellular signaling mechanism essential for proper embryonic development in all metazoan

organisms in the animal kingdom. According to the KEGG annotation, the Notch proteins (Notch1-Notch4 in vertebrates) are single-pass receptors that are activated by the Delta (or Delta-like) and Jagged/Serrate families of membrane-bound ligands. They are transported to the plasma membrane as cleaved, but otherwise intact polypeptides. Interaction with ligand leads to two additional proteolytic cleavages that liberate the Notch intracellular domain (NICD) from the plasma membrane. The NICD translocates to the nucleus, where it forms a complex with the DNA binding protein CSL, displacing a histone deacetylase (HDAc)-co-repressor (CoR) complex from CSL. Components of an activation complex, such as MAML1 and histone acetyltransferases (HATs), are recruited to the NICD-CSL complex, leading to the transcriptional activation of Notch target genes.

Unsupervised Cluster Analysis

The signature gene set for Notch signaling pathway was retrieved from MSigDB (Subramanian, et al., 2005). The gene set included a total of 47 genes which were shown to be integral genes in the Notch signaling pathway. A hierarchical clustering of both the DNA methylation and gene expression intensities revealed that most of the genes were similarly expressed in tumors and OCICs. The OCIC samples clustered with tumor samples with a very significant correlation of greater than 0.8 which signifies the close similarities in the expression patterns in the two phenotypes. The methylation patterns observed in this signature is consistent with previous analysis suggesting more methylation observed in genes belonging to OCIC compared to tumor and normal OSE, whereas the later phenotypes show less differences in the methylation patterns. OCIC on the other hand showed great divergence in the methylation patterns compared to the other two phenotypes.

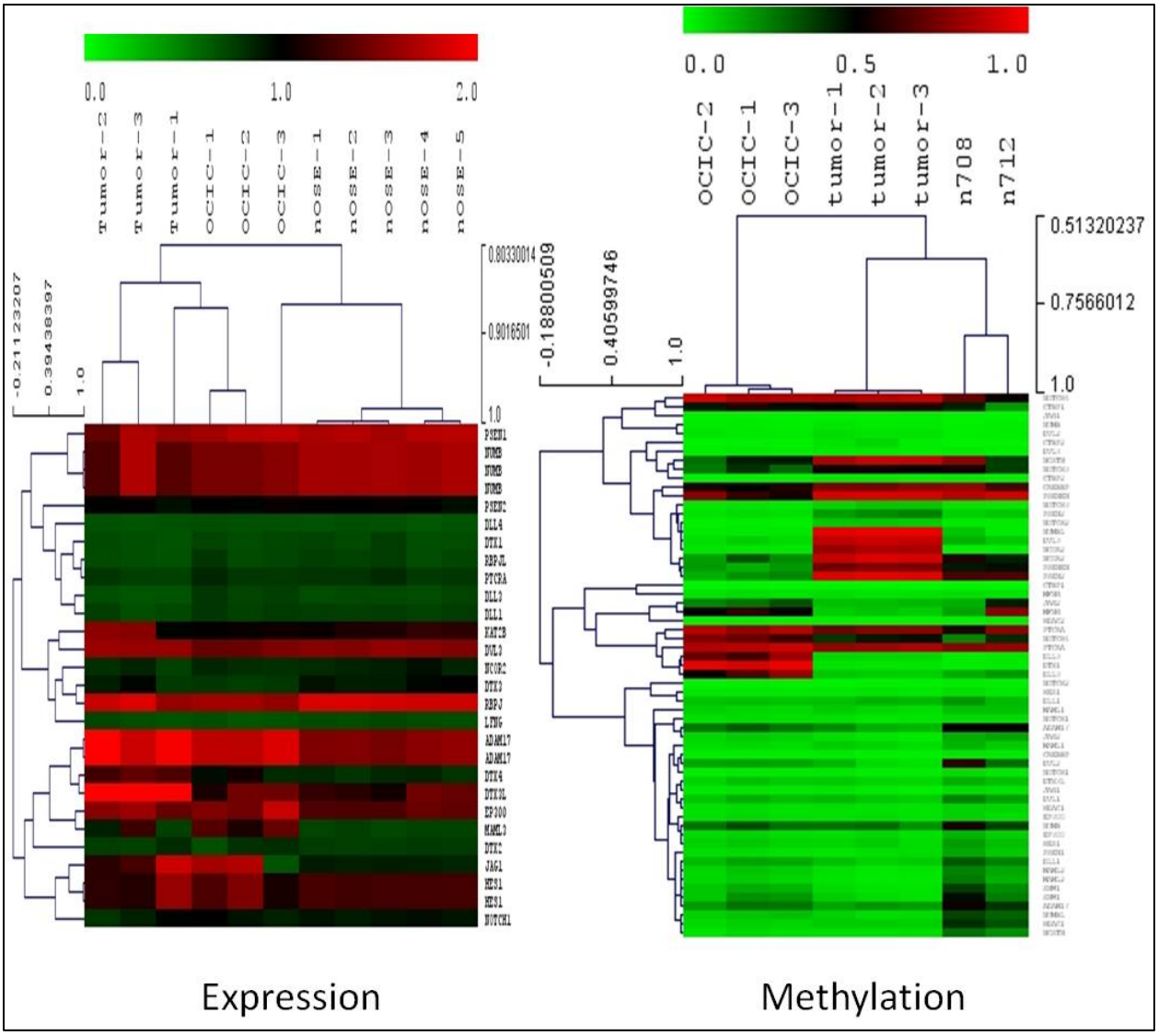

Figure 28: Unsupervised clustering of the genes from Notch signaling pathway from KEGG.

The expression patterns of Notch signaling pathway were further studied using differential expression in the genes belonging to tumor and OCIC phenotypes compared to normal OSE. The genes were selected based on the differential expression criteria (explained in methods section) and also the fold change. The details of this analysis can be seen in Figures 29 and 30. Consistent with the unsupervised analysis, the differential

expression patterns showed similar gene expression patterns in the Tumor and OCIC cells with some differences in the expression of few up-stream genes.

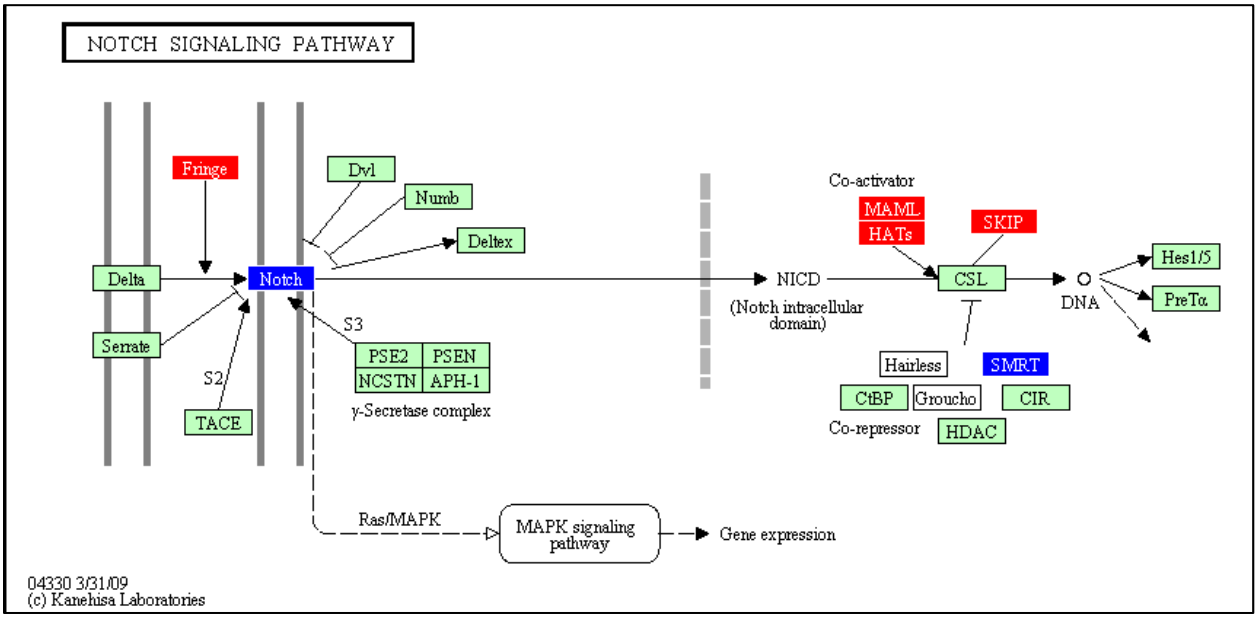

Figure 29: Differential analysis for Notch signaling pathway for OCIC compared to normal OSE.

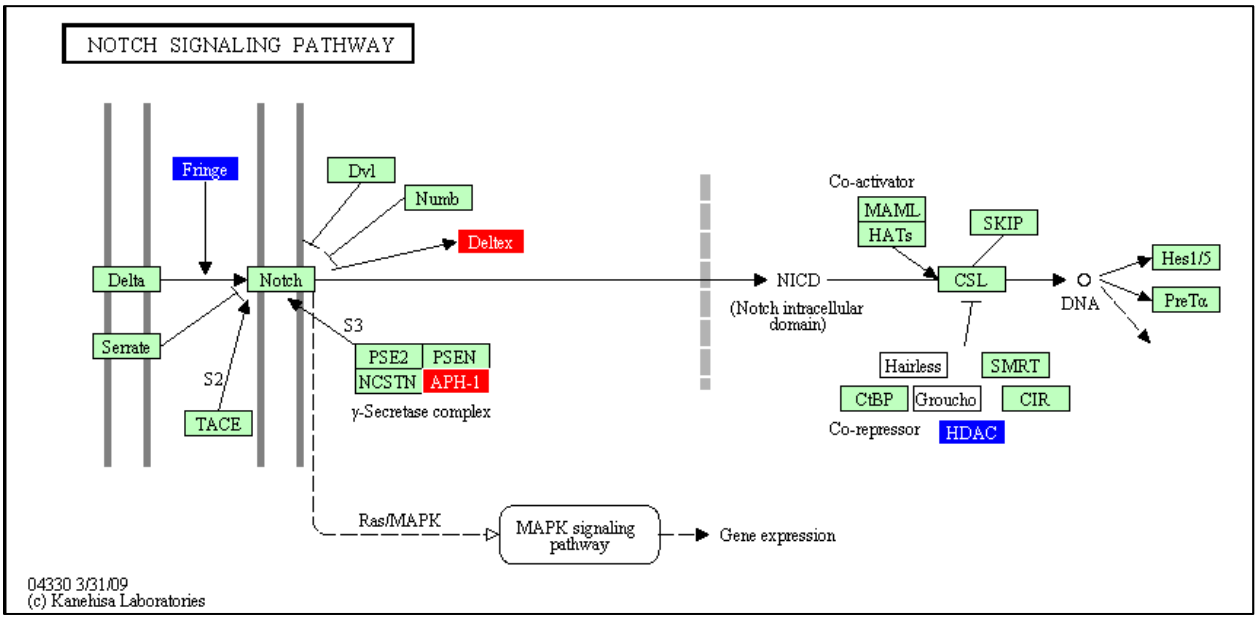

Figure 30: Differential analysis for Notch signaling pathway for Tumor compared to normal OSE.

It was seen that the Notch gene which regulates normal gene regulation through MAPK signaling pathway was seen to be significantly down-regulated in OCIC which was not evident in tumor as the change in the gene expression for tumor was not significant for Notch gene. On the other hand, the Fringe genes which are important for the working of the Notch signaling was seen to be upregulated in OCIC but down-regulated in Tumors. As the OCICs cluster closely with tumors, and as the notch signaling is known to be responsible for intercellular communication which is usually disrupted in cancer, the analysis suggests the similar disruption in the cellular communication can be prevalent in OCICs.

Comparison with Fallopian Tube Datasets

As explained earlier, the public domain fallopian tube dataset included three main phenotypes, Fallopian Tube Epithelial Normal Cells (FTEn), Fallopian Tube Serous Carcinoma (FTSerCa) and Ovarian Cancer Serous Carcinoma (OVSerCa). The tumors from the in-house dataset were matched and integrated with the OVSerCa phenotype as stated earlier. The integrated datasets were further used for unsupervised hierarchical clustering and differential expression analysis which will be discussed here. The integrated dataset was tested with the similar gene signatures determined in the earlier analysis with intent to compare the expression patterns of OCICs as precursor of Tumors with Fallopian Tube cells and in the process speculate on the genesis of ovarian cancer based on the similarities in expression patterns of OCIC with either normal OSE or Fallopian tube epithelium.

Epithelial – Mesenchymal Transition Signature

The signature for EMT which was used earlier with two datasets; up-regulated and down-regulated was again used for this analysis. Out of a total of 91 up-regulated

signature genes, the integrated dataset contained information for 63 genes, whereas out of the 159 down-regulated genes, the data was available for 82 genes.

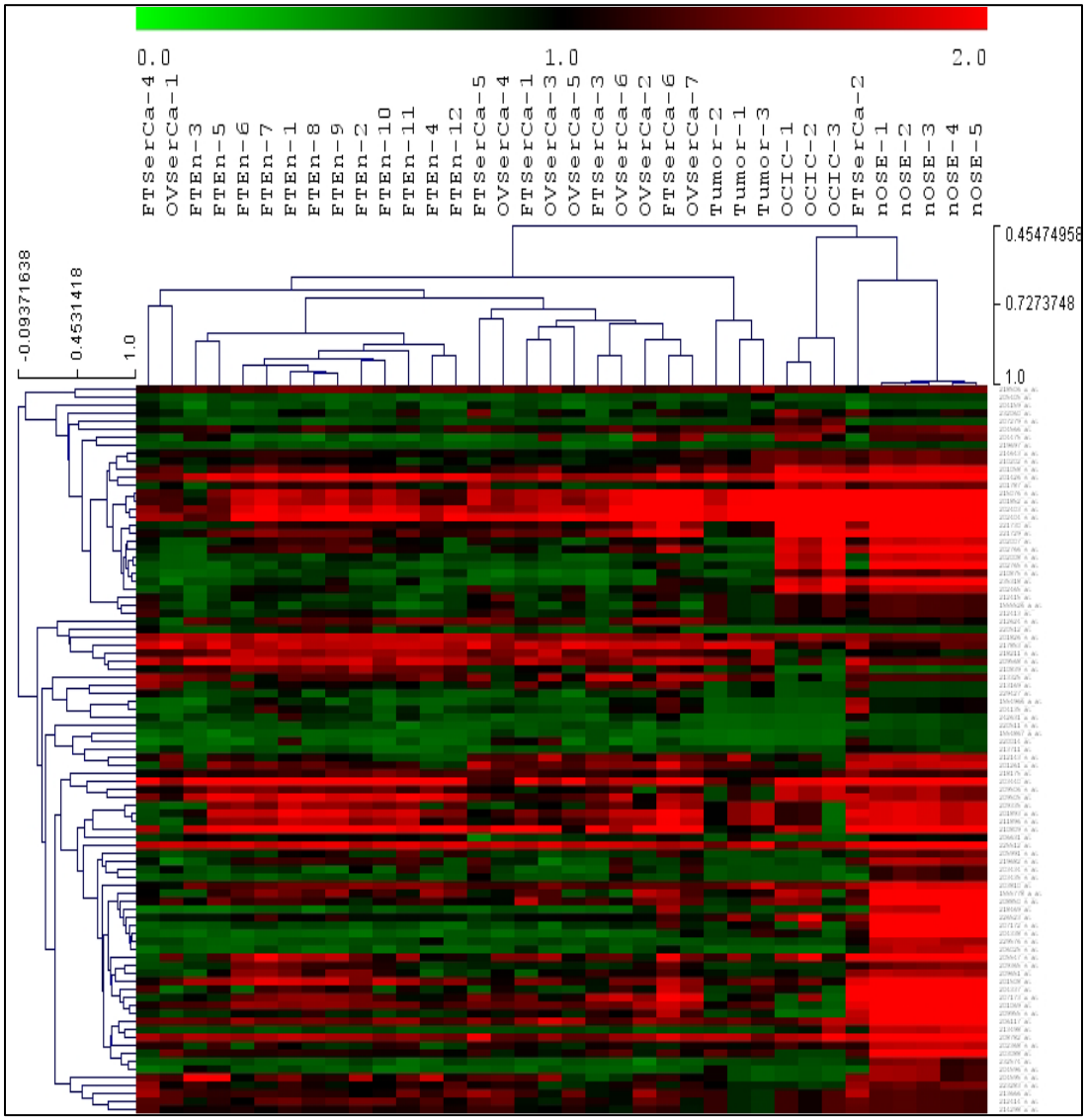

Figure 31: Unsupervised clustering of the up-regulated genes from the EMT signature.

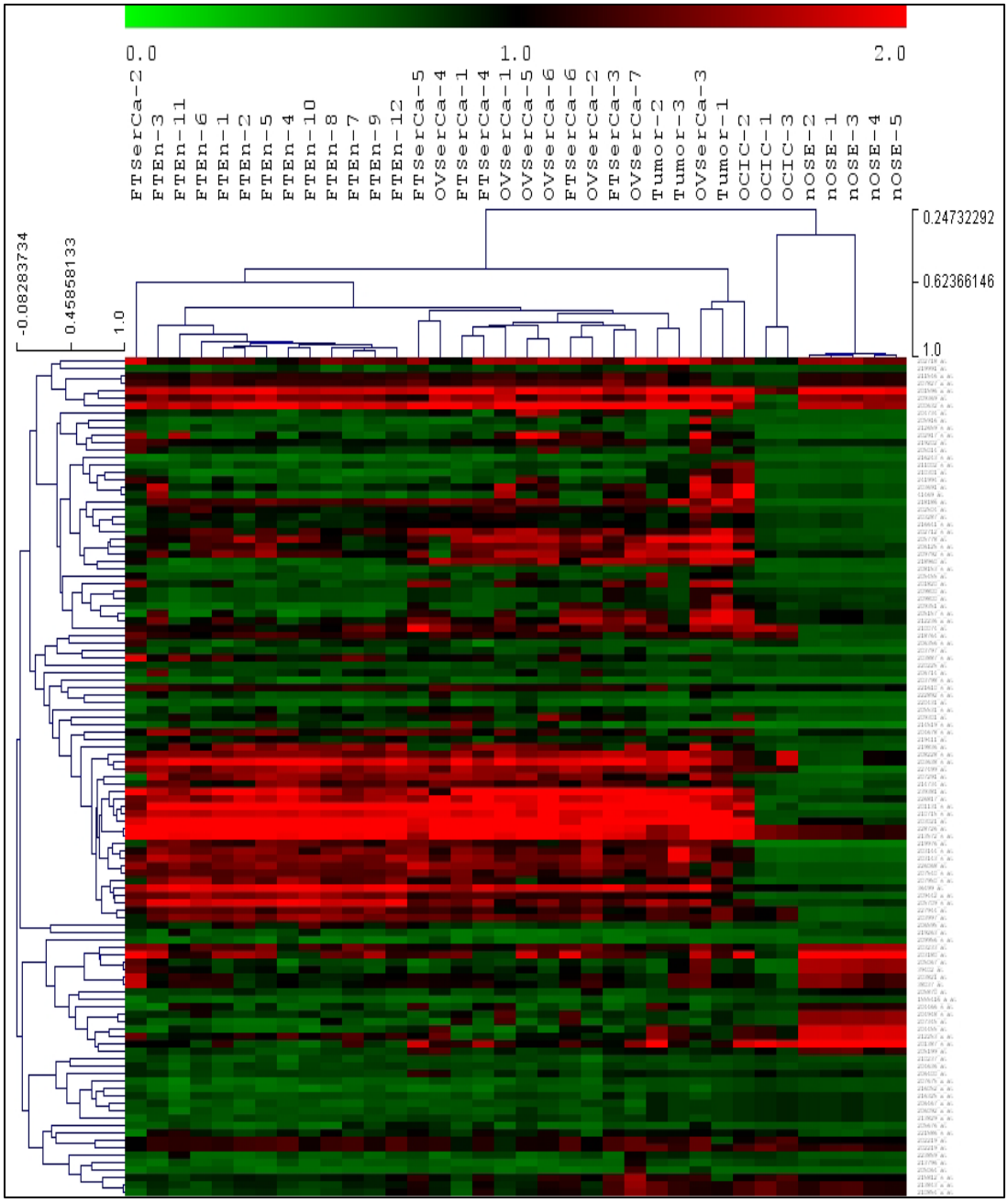

Figure 32: Unsupervised clustering of the down-regulated genes from the EMT signature.

As observed in the earlier analysis of EMT signature, it was seen that OCICs show a mesenchymal phenotype where as Tumors show an epithelial phenotype. This can be seen from the above clustering patterns, the OCICs cluster closely with nOSE cells

where most of the down regulated genes in the signature are seen to be down regulated in these two phenotypes and the upregulated genes in the signature seen to be upregulated. This is consistent with the fact that biologically normal OSE cells are known to be mesenchymal showing the EMT process to be active. On the other hand, the Tumors cluster closely with the normal Fallopian tube epithelial cells which show all the epithelial properties where EMT processes are inactive. The tumors thus can be said to show an epithelial phenotype. Further, the normal OSE and Fallopian tube cells are highly divergent form each other clustering significantly away from each other, showing distinct gene expression patterns.

TGF-beta Signaling Pathway Signature

As stated earlier, TGF-beta signaling regulated wide range of cellular functions such as cell proliferation, apoptosis, cell differentiation and migration. In order to study the similarities between tumors and Fallopian tube epithelial cells, we studied the TGF-beta signature from MSigDB which provides all the genes in the TGF-beta pathway annotated in KEGG database. The signature included 86 genes which were known to play an important part in the TGF-beta signaling pathway. Out of these genes, intensity information was available for 47 genes in the integrated dataset.

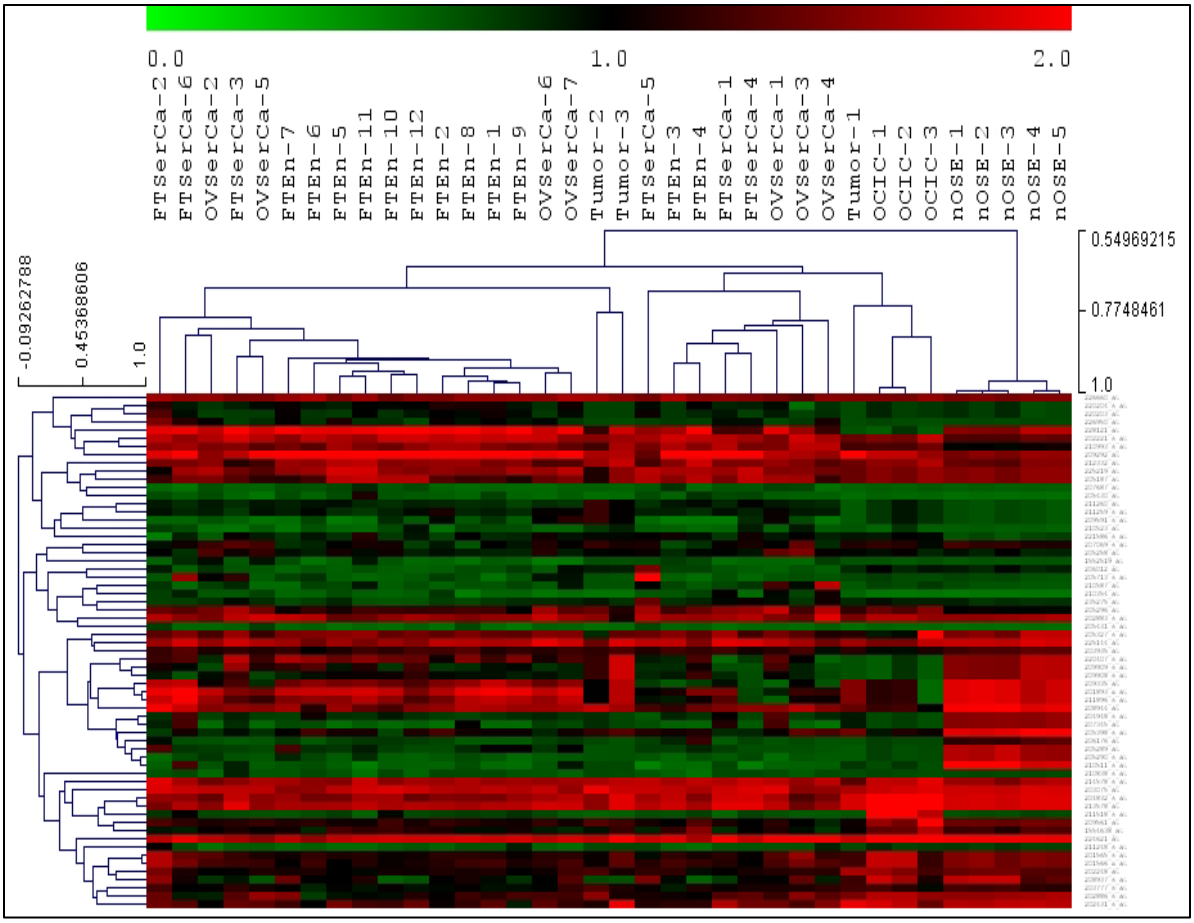

Figure 33: Unsupervised clustering of the genes from the TGF-beta signature.

Consistent with the TGF-beta signature studied earlier, we observed that the OCICs cluster together with tumors and in this case also with normal fallopian tube epithelial samples. This suggests that the TGF-beta signaling patterns observed in tumors and OCICs were more similar to those in fallopian epithelial cells than the ovarian surface epithelial cells. The expression patterns of normal OSE were again vastly different to normal Fallopian tube epithelial calls. The signature was further studied by performing differential analysis between the Tumors and OCICs compared to the fallopian tube epithelial cells. The details of this analysis for the TGF-beta signaling pathway are illustrated in the Figures 35 and 35.

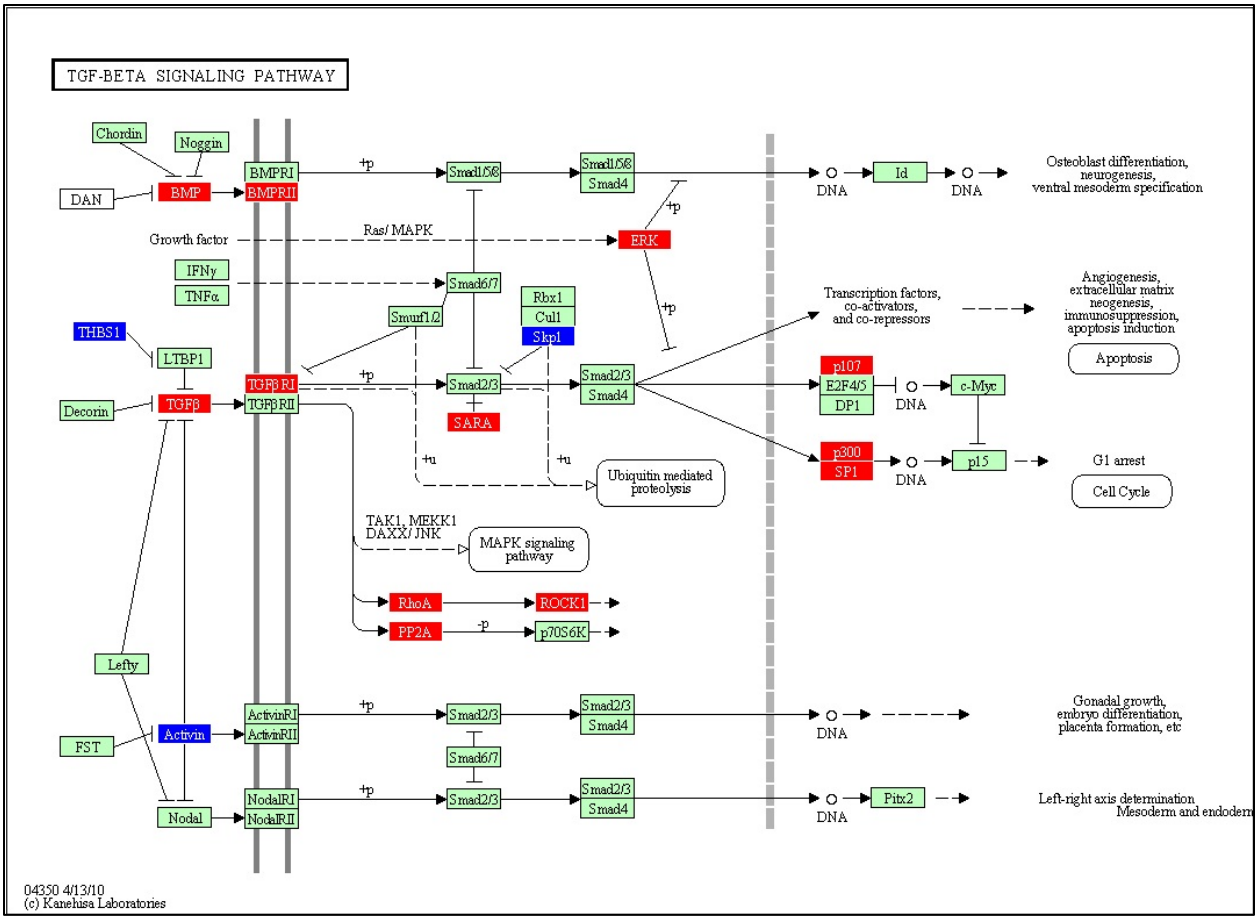

Figure 34: Differential analysis for TGF-beta signaling pathway for OCIC compared to normal Fallopian Tube epithelial cells.

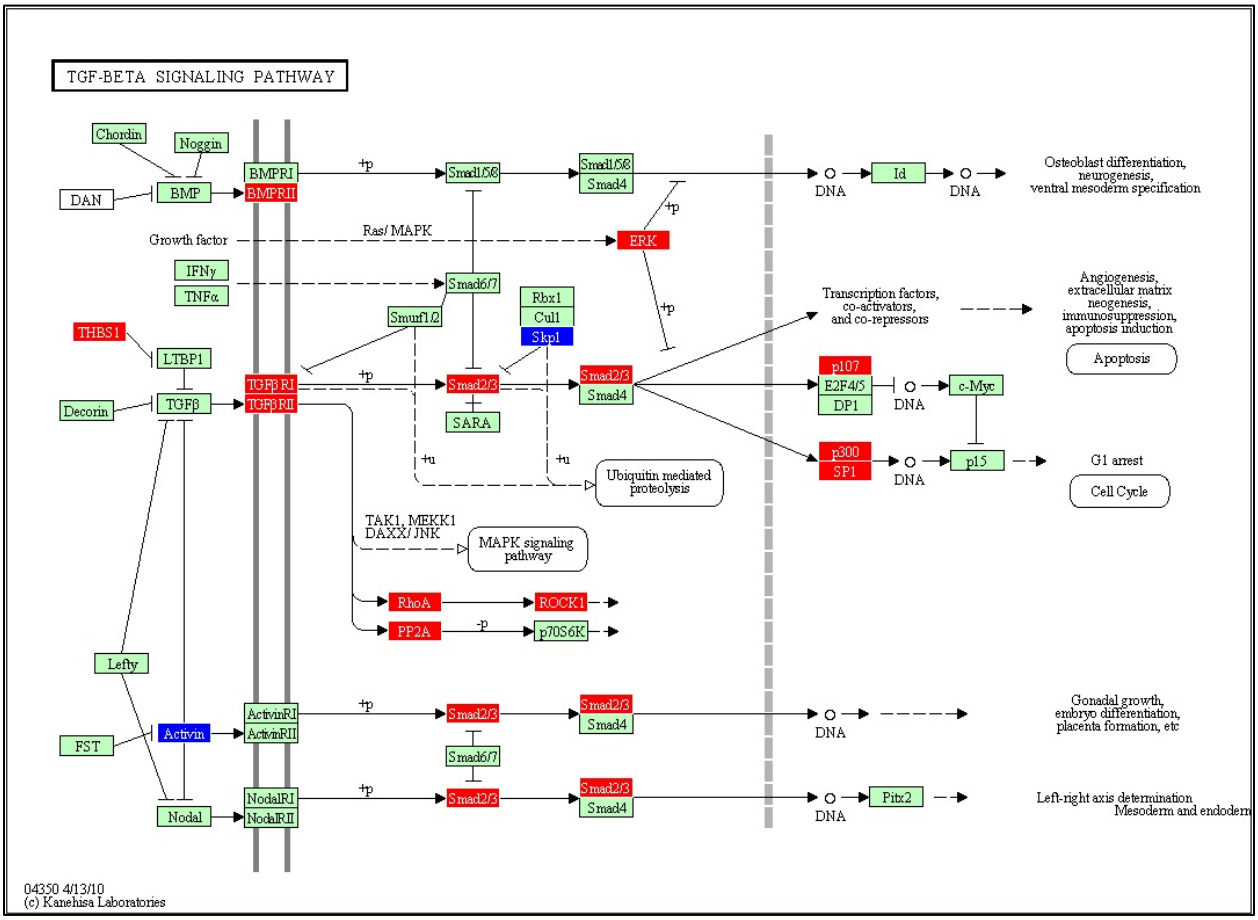

Figure 35: Differential analysis for TGF-beta signaling pathway for Tumor compared to normal Fallopian Tube epithelial cells.

It was observed that most of the genes regulating the TGF-beta signaling pathway were upregulated in both the Tumor and OCICs compared to fallopian tube epithelium. This analysis confirmed that the genes involved in the TGF-beta pathway showed similar expression patterns for both the OCICs and Tumors compared to the fallopian tube epithelial cells showing an upregulation of the TGF-beta signaling pathway.

Wnt signaling pathway

The signature genes representing the Wnt pathway that were explained in the earlier analysis were again used in this analysis for the comparison of OCICs with the

fallopian tube epithelial cells. As stated earlier, Wnt pathway regulates many basic developmental processes like cell-fate specification, progenitor cell proliferation and the control of asymmetric cell division, the inactivation of this pathway is a signature of the tumor phenotypes which lead to asymmetric regulation of the above pathways leading to cancer. The comparison of this signature with the Fallopian tube samples will indicate their similarities to the epithelial phenotype. The clustering of OCIC and Tumors compared to the fallopian tube and ovarian surface epithelial cells can be seen in Figure-36. It was seen that the expression data was available for only 75 genes in the integrated data out of the total 151 genes in the signature.

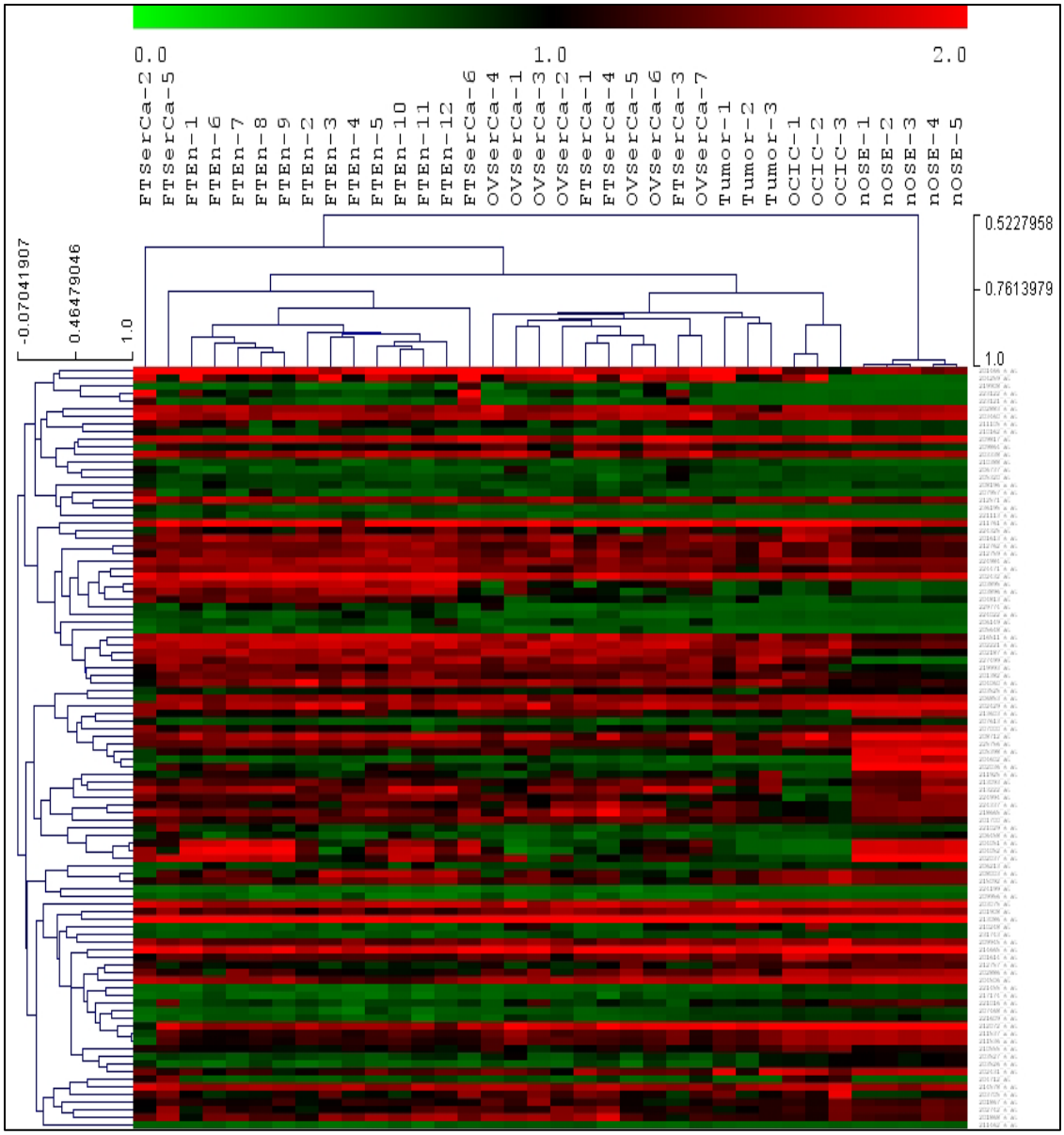

Figure 36: Unsupervised clustering of the genes from the WNT Signature.

The clustering patterns of OCICs were shown to be more similar to the tumors where, both tumors and OCICs clustered with epithelial cells compared to the normal OSE cells. This signature thus indicates the similarities in the Tumors and epithelial cells with the OCICs. This was further tested using differential expression analysis on this signature. The differentially expressed genes between the tumor and OCIC compared to

Fallopian epithelial cells can be seen in Figures 37 and 38. The comparison will indicate the similarities and differences of the expression patterns in OCICs and Tumors.

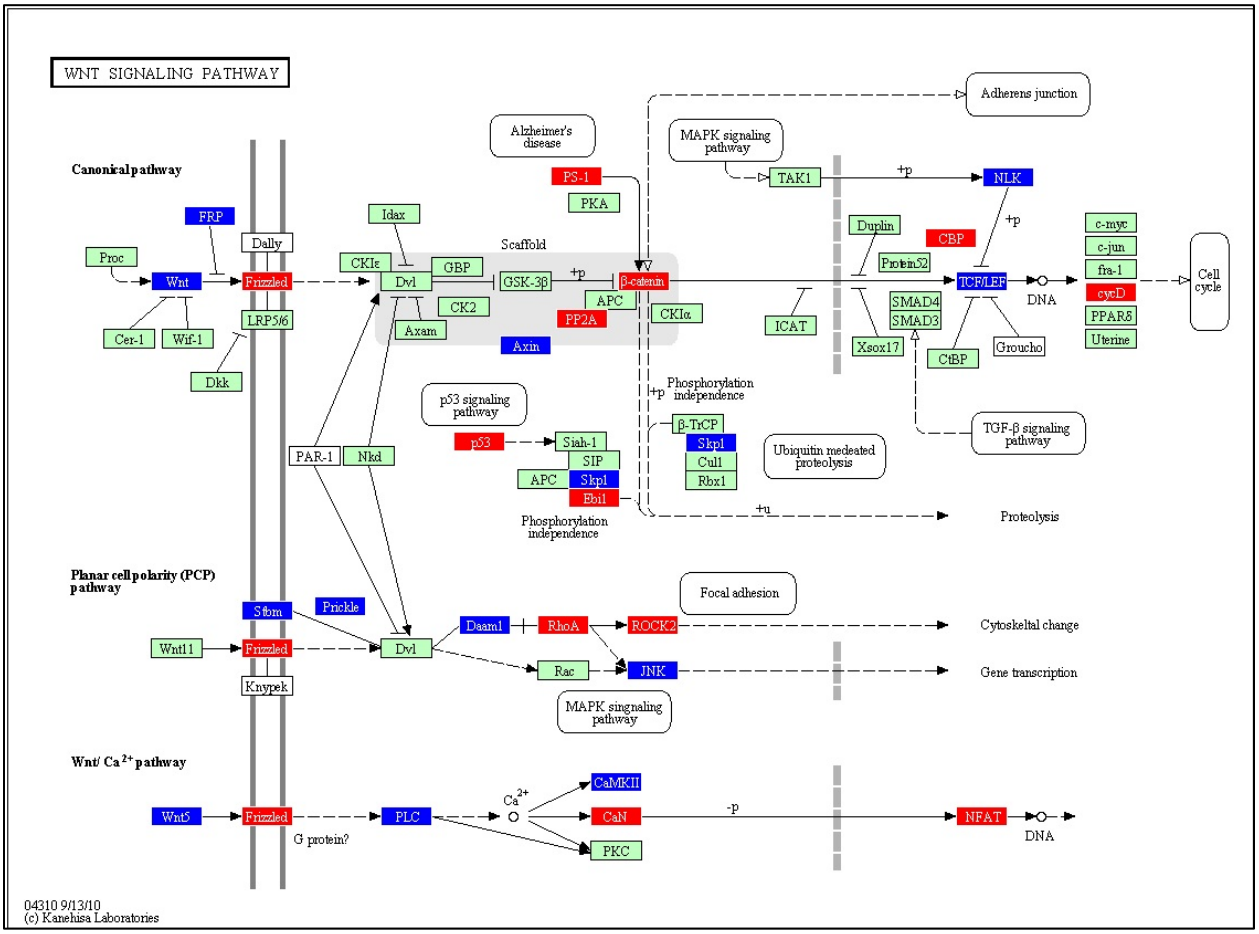

Figure 37: Differential analysis for Wnt signaling pathway for OCICs compared to normal Fallopian Tube epithelial cells.

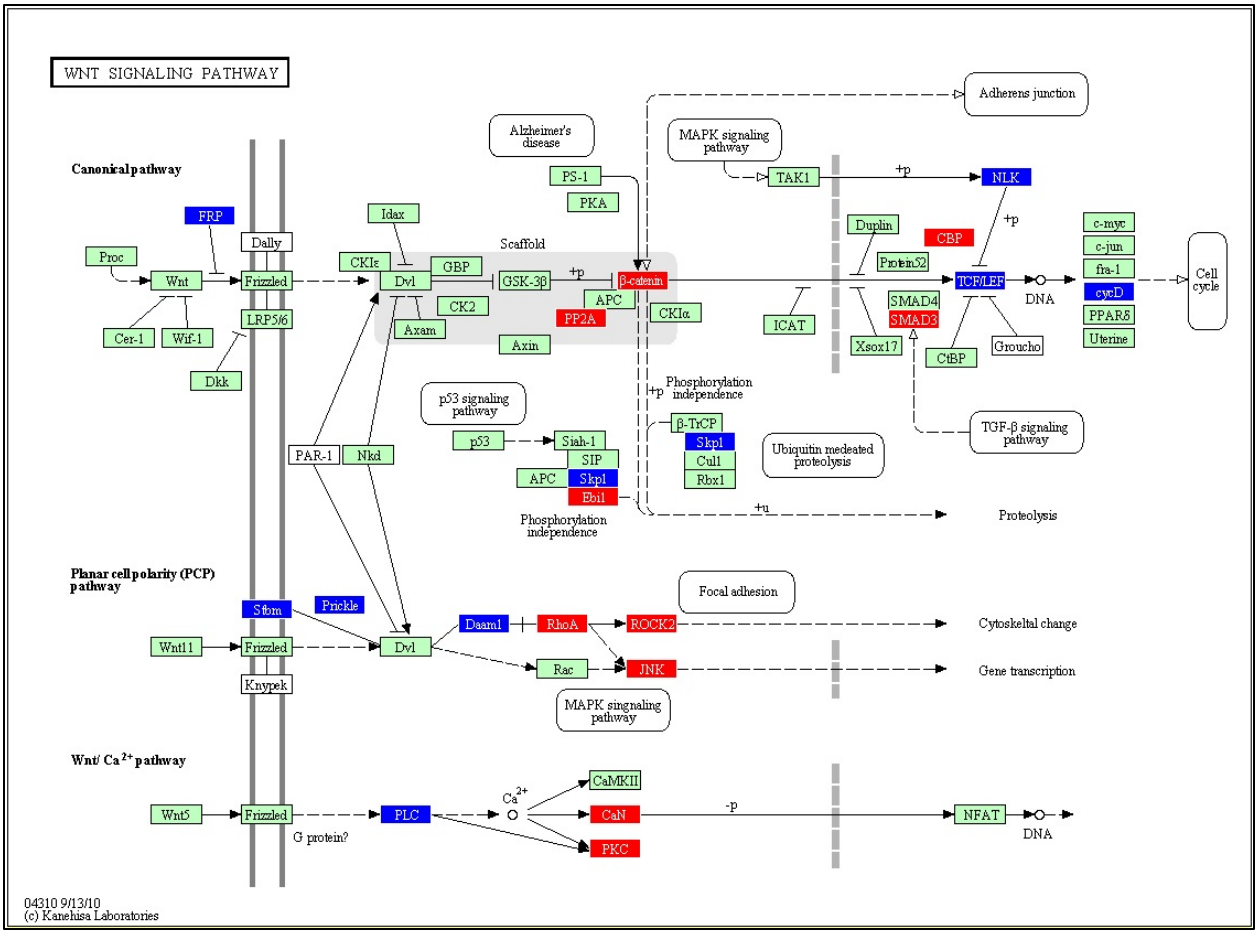

Figure 38: Differential analysis for Wnt signaling pathway for Tumor compared to normal Fallopian Tube epithelial cells.

The differential expression analysis revealed that the expression patterns for OCICs and Tumors are very similar compared to Fallopian Tube epithelial cells. It was seen that Wnt genes were mostly down-regulated in OCICs and Tumors which is similar to the results for the comparison with the normal OSE cells. It was also seen that consistent with the previous analysis, some of the cell cycle regulating genes were down regulated in tumor but were up-regulated in OCICs. Overall, the expression patterns found in the differential comparisons of the comparisons with the Fallopian tube and normal OSE showed similar expression patterns for both phenotypes.

Notch Signaling Pathway

The signature genes representing the Notch signaling pathway that were explained in the earlier analysis were again used in this analysis for the comparison with the fallopian tube epithelial cells. As stated, Notch pathway regulates many important processes like inter-cell communication, cell differentiation and cell fate decisions; the inactivation of this pathway is a signature of the tumor phenotypes which lead to asymmetric regulation of the above pathways leading to cancer. The comparison of this signature with the Fallopian tube samples will indicate their similarities to the epithelial phenotype. The clustering of OCIC and Tumors compared to the fallopian tube and ovarian surface epithelial cells can be seen in Figure-39. It was seen that the expression data was available for only 24 genes in the integrated data out of the total 47 genes in the signature.

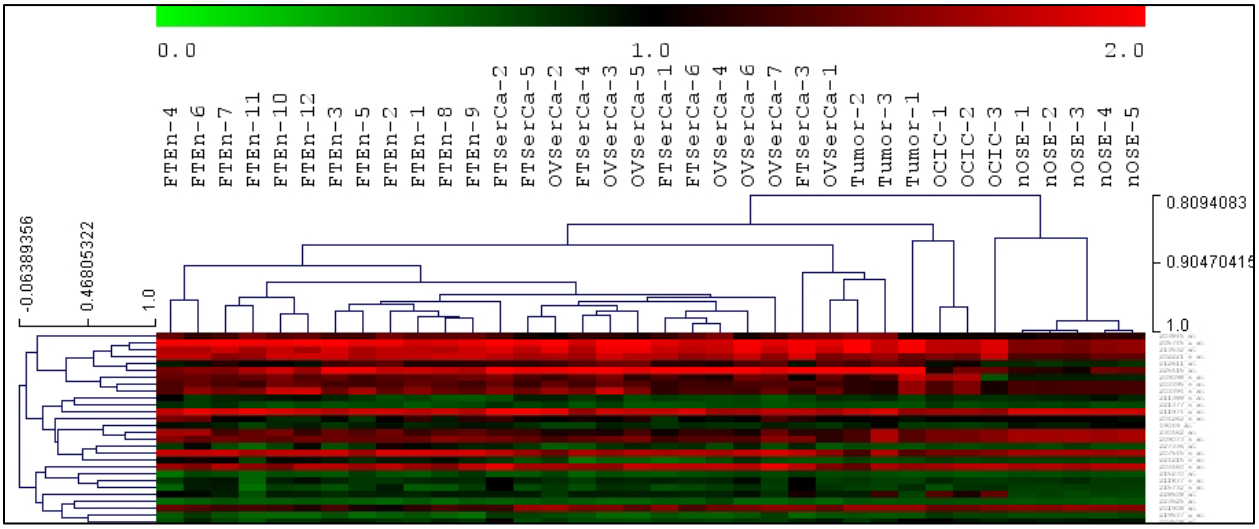

Figure 39: Unsupervised clustering of the genes from the Notch Signature.

Even though the numbers of genes were less in this signature, the clustering patterns of OCICs were shown to be more similar to the tumors where, both tumors and OCICs clustered with epithelial cells compared to the normal OSE cells with a greater

than 0.8 correlation. This signature thus indicates the similarities in the Tumors and epithelial cells with the OCICs. This was further tested using differential expression analysis on this signature. The differentially expressed genes between the tumor and OCIC compared to Fallopian epithelial cells can be seen in Figures 40 and 41. The comparison will also indicate the similarities and differences of the expression patterns in OCICs and Tumors.

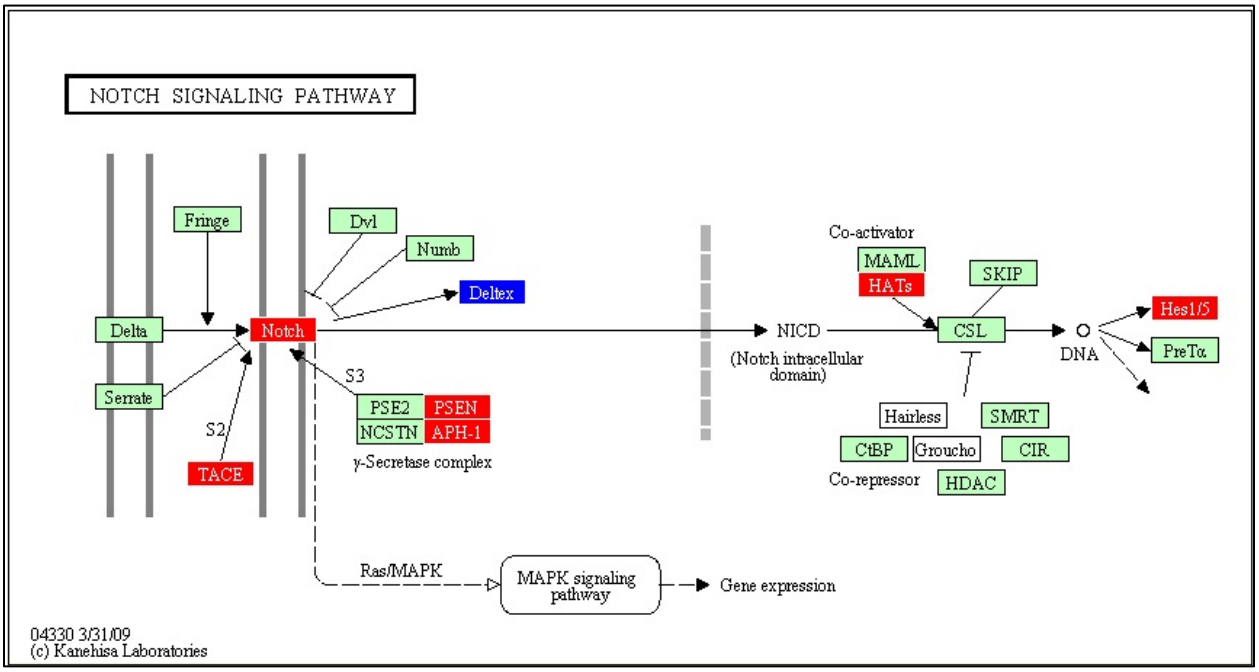

Figure 40: Differential analysis for Notch signaling pathway for OCICs compared to normal Fallopian Tube epithelial cells.

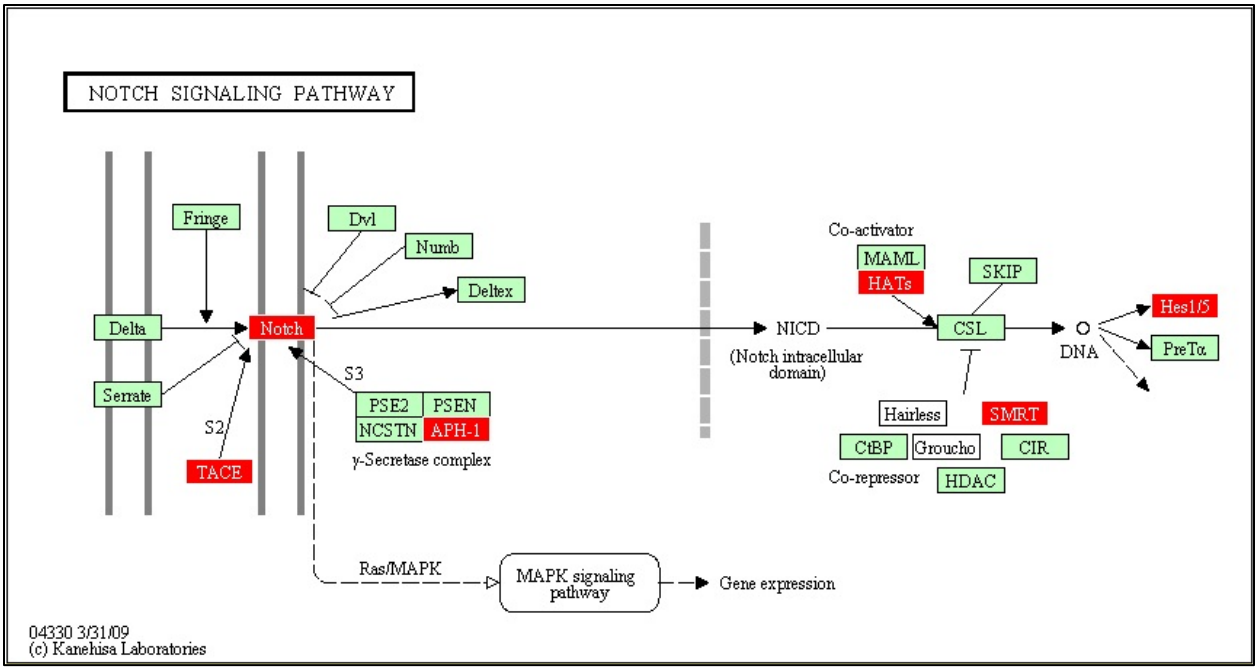

Figure 41: Differential analysis for Notch signaling pathway for Tumor compared to normal Fallopian Tube epithelial cells.

The differential expression analysis revealed that the expression patterns for OCICs and Tumors are very similar compared to Fallopian Tube epithelial cells. It was seen that Notch genes were mostly up-regulated in OCICs and Tumors which is contrary to the results for the comparison with the normal OSE cells where Notch was down-regulated. These expression patterns found in the differential comparisons with the Fallopian tube and normal OSE suggest that the Notch pathway is active in tumors and OCICs with respect to the epithelial cells but its activation is not at the levels of the mesenchymal cells, thus suggesting them to be intermediates between the mesenchymal and epithelial phenotypes.

Chapter 6 DISCUSSION

Role of ovarian cancer initiating cells in cancer development and progression

Recently a subpopulation of tumor cells known as ovarian cancer initiating cells (OCICs) have been shown to be the cells that propagate the tumor phenotype in ovarian cancer. The studies have showed that a very small population (100 cells) of these cells are sufficient to induce a tumor phenotype, while a large quantity of cells ($5 \text{ X } 10^5$) cell are required to induce such a phenotype in xenotransplanted mouse. In this thesis we studied the functional changes in genes present in the OCIC phenotype which were important for such efficient propagation of cancers. To enable such analysis, we generated mRNA expression and DNA methylation profiles of OCICs and compared them with those of Tumor and normal ovarian surface epithelial cells. From these analyses and also from various studies in literature which analyzed similar cells in other cancers, we determined four pathways which regulated most of the observed changes and were predicted to be important factors in distinguishing the OCICs from Tumors and normal cells. We identified gene signatures from the literature which were regulated by these pathways and performed unsupervised cluster analysis on these signatures in order to determine the similarities of OCICs with respect to tumor and normal samples.

These analyses gave us a better understanding of the OCICs. We determined that even though OCICs have been isolated from the Tumor cells, they show distinct gene expression and DNA methylation profiles. These cells were found to be an intermediate subtype between the mesenchymal and epithelial phenotypes. They were shown to retain some of the EMT signature gene profiles form normal ovarian surface epithelial cells which showed OCICs to be mesenchymal by clustering them with the mesenchymal

normal OSE cells. On the other hand, for the cancer related pathways like TGF-beta, Wnt and Notch, the OCIC cells shared expression profiles with tumor cells, thus making them altered cells similar to tumors. The OCICs consistently showed methylation profiles distinct from the tumors and normal OSE cells for all the signatures. They also showed significantly more genes to be hypermethylated than the other two phenotypes for all the signatures thus indicating a scope of using the genes from these signatures as DNA methylation markers.

The down-regulation of the tumor suppressor pathways in OCICs could explain their ability to propagate tumors. On the other hand, the mesenchymal properties allow them to migrate and form metastases. These properties are symbolic of a cancer phenotype. On the contrary, even though the tumor cells showed down regulation of tumor suppressor pathways which can enable them to propagate the tumors, they showed an epithelial phenotype. This epithelial phenotype of the tumors eliminates the possibility of these cells to migrate and in turn forming metastases. The analysis suggests that the OCICs are the small populations of tumors that retain the mesenchymal or migratory characteristics while the other tumor cells lack this property.

Insight in the genesis of ovarian cancer

We believed that the OCICs can be used as indicators towards the genesis and progression of early events in the ovarian cancers. In light of this, we considered two hypotheses which are currently addressing the genesis of ovarian cancer. The first hypothesis proposed ovarian surface epithelial cells to be cells of origin of the ovarian cancer while the other proposed the Fallopian Tube cells to be contributing the cell of origin for these cancers. It is also believed that these two cells can be reciprocal cells of

origin for the cancer phenotype (Landen, et al., 2008). In order to test these hypotheses, we integrated the in-house dataset with a public domain Fallopian tube gene expression data.

The comparison of the OCICs with the fallopian tube cells revealed that as the tumors become more epithelial during their proliferation, most of the tumor related pathways show expression patterns which are similar to the epithelial cells for tumors and also for OCICs. The exception to this pattern is the EMT signature which indicated that the OCICs are closer to the normal OSE cells showing a mesenchymal phenotype for both. The interesting fact consistent for all the comparisons was that the normal OSE were shown to be highly divergent form fallopian epithelial cells for all the signatures.

The analysis thus gives an indication of normal OSE cells to be the cells of origin as OCIC being the precursors of tumors, can be thought to retain the mesenchymal properties from OSE cells. It also suggests that the fallopian tube epithelial cells and normal OSE have completely different expression profiles and cannot be thought as reciprocal cells of origin, in fact the gene expression patterns suggest that they are mutually exclusive cells. Also, the reason for the OCICs clustering with the tumors and also with fallopian epithelial cells for the cancer related signatures can be due to the fact that the tumors are known to show an epithelial phenotype after they proliferate form the OCICs and their similarities in the expression patterns with OCICs can be indicative of the initiation of the development of the epithelial phenotype in the later.

Proposed Model of Ovarian Carcinogenesis

There has been a constant rise in the knowledge of the early events in ovarian cancer, which has led to a better understanding of the genesis and progression of tumors

in ovarian cancer. However, we propose that in a comprehensive model of ovarian cancer, the ovarian cancer initiating cell (OCIC) is the vital component which originated due to prolonged physicochemical effects on ovarian surface epithelial cells. These ovarian surface epithelial cells known to possess mesenchymal properties (to carry out postovulatory functions) conserve the mesenchymal property after their transformation to the malignant ovarian cancer initiating cells. The malignant properties of these cells are reflected by the down-regulation of various tumor suppressor pathways like TGF-beta, Wnt and Notch signaling pathway. During the transformation of ovarian surface epithelium to OCICs, many genes belonging to the tumor suppressor pathways undergo DNA methylation enabling the down-regulation of these pathways. These malignant cancer initiating cells proliferate into tumors which are differentiated cells possessing most of the epithelial properties and conserve the malignant properties i.e. down-regulation of tumor suppressor pathways. The methylation subsequently is removed to some extent during the proliferation and differentiation of OCICs into tumor cells. The OCICs having inherited the mesenchymal properties will eventually migrate to distant sites and proliferate thus creating metastatic tumors.

Future implications

It is believed that standard chemotherapy procedures fail to target the tumor progenitor cells like ovarian cancer initiating cells. The main reason for this is that they express normal stem cell phenotypes like low mitotic index, enhanced DNA repair and expression of membrane efflux transporters (Zhang, et al., 2008). The potential DNA methylation markers identified in this thesis by the hypermethylation observed in the genes from the signature pathways like EMT, TGF-beta, Wnt and Notch can be used to

target and eliminate OCICs in the future therapies. The further characterization of these progenitors and their comprehensive studies for chromatin or Histone modifications and changes in miRNA expression patterns will provide us with a greater understanding in the alterations of these key pathways in ovarian cancer and will allow development of better methods for early detection of this highly elusive disease.

Chapter 7 REFERENCES

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A, 100*(7), 3983-3988.

Auersperg, N., Wong, A. S., Choi, K. C., Kang, S. K., & Leung, P. C. (2001). Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr Rev, 22*(2), 255-288.

Baba, T., Convery, P. A., Matsumura, N., Whitaker, R. S., Kondoh, E., Perry, T., et al. (2009). Epigenetic regulation of CD133 and tumorigenicity of CD133+ ovarian cancer cells. *Oncogene, 28*(2), 209-218.

Balch, C., Montgomery, J. S., Paik, H. I., Kim, S., Huang, T. H., & Nephew, K. P. (2005). New anti-cancer strategies: epigenetic therapies and biomarkers. *Front Biosci, 10*, 1897-1931.

Bapat, S. A., Mali, A. M., Koppikar, C. B., & Kurrey, N. K. (2005). Stem and progenitor-like cells contribute to the aggressive behavior of human epithelial ovarian cancer. *Cancer Res, 65*(8), 3025-3029.

Bartolomei, M. S., & Tilghman, S. M. (1997). Genomic imprinting in mammals. *Annu Rev Genet, 31*, 493-525.

Baylin, S. B. (2005). DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol, 2 Suppl 1*, S4-11.

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol, 203*(4), 971-983.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E. W., Wu, B., et al. (2006). High-throughput DNA methylation profiling using universal bead arrays. *Genome Res, 16*(3), 383-393.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev, 16*(1), 6-21.

Bird, A. P., & Wolffe, A. P. (1999). Methylation-induced repression--belts, braces, and chromatin. *Cell, 99*(5), 451-454.

Bowtell, D. D. (2010). The genesis and evolution of high-grade serous ovarian cancer. *Nat Rev Cancer, 10*(11), 803-808.

Clarke, M. F., Dick, J. E., Dirks, P. B., Eaves, C. J., Jamieson, C. H., Jones, D. L., et al. (2006). Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res, 66*(19), 9339-9344.

Collins, A. T., Berry, P. A., Hyde, C., Stower, M. J., & Maitland, N. J. (2005). Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res, 65*(23), 10946-10951.

Das, P. M., & Singal, R. (2004). DNA methylation and cancer. *J Clin Oncol, 22*(22), 4632-4642.

Du, P., Kibbe, W. A., & Lin, S. M. (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics, 24*(13), 1547-1548.

Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., et al. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics, 11*, 587.

Dubeau, L. (2008). The cell of origin of ovarian epithelial tumours. *Lancet Oncol, 9*(12), 1191-1197.

Fang, D., Nguyen, T. K., Leishear, K., Finko, R., Kulp, A. N., Hotz, S., et al. (2005). A tumorigenic subpopulation with stem cell properties in melanomas. *Cancer Res, 65*(20), 9328-9337.

Feinberg, A. P., & Tycko, B. (2004). The history of cancer epigenetics. *Nat Rev Cancer, 4*(2), 143-153.

Fojo, T., & Bates, S. (2003). Strategies for reversing drug resistance. *Oncogene, 22*(47), 7512-7523.

Fraser, M., Leung, B., Jahani-Asl, A., Yan, X., Thompson, W. E., & Tsang, B. K. (2003). Chemoresistance in human ovarian cancer: the role of apoptotic regulators. *Reprod Biol Endocrinol, 1*, 66.

Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., et al. (2004). Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res, 64*(19), 7011-7021.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., et al. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science, 311*(5759), 395-398.

Grossman, D., & Altieri, D. C. (2001). Drug resistance in melanoma: mechanisms, apoptosis, and new potential therapeutic targets. *Cancer Metastasis Rev, 20*(1-2), 3-11.

Gupta, P. B., Chaffer, C. L., & Weinberg, R. A. (2009). Cancer stem cells: mirage or reality? *Nat Med, 15*(9), 1010-1012.

Herman, J. G. (1999). Hypermethylation of tumor suppressor genes in cancer. *Semin Cancer Biol, 9*(5), 359-367.

Hope, K. J., Jin, L., & Dick, J. E. (2004). Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat Immunol, 5*(7), 738-743.

Houshdaran, S., Hawley, S., Palmer, C., Campan, M., Olsen, M. N., Ventura, A. P., et al. (2010). DNA methylation profiles of ovarian epithelial carcinoma tumors and cell lines. *PLoS One, 5*(2), e9359.

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res, 37*(1), 1-13.

Integrated genomic analyses of ovarian carcinoma. (2011). *Nature, 474*(7353), 609-615.

Jaenisch, R., Beard, C., Lee, J., Marahrens, Y., & Panning, B. (1998). Mammalian X chromosome inactivation. *Novartis Found Symp, 214*, 200-209; discussion 209-213, 228-232.

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet, 3*(6), 415-428.

Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nat Genet, 21*(2), 163-167.

Knoblich, J. A. (2008). Mechanisms of asymmetric stem cell division. *Cell, 132*(4), 583-597.

Landen, C. N., Jr., Birrer, M. J., & Sood, A. K. (2008). Early events in the pathogenesis of epithelial ovarian cancer. *J Clin Oncol, 26*(6), 995-1005.

Lehrmann, H., Pritchard, L. L., & Harel-Bellan, A. (2002). Histone acetyltransferases and deacetylases in the control of cell proliferation and differentiation. *Adv Cancer Res, 86*, 41-65.

Levanon, K., Crum, C., & Drapkin, R. (2008). New insights into the pathogenesis of serous ovarian cancer and its clinical impact. *J Clin Oncol, 26*(32), 5284-5293.

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell, 69*(6), 915-926.

Liu, L., Wylie, R. C., Andrews, L. G., & Tollefsbol, T. O. (2003). Aging, cancer and nutrition: the DNA methylation connection. *Mech Ageing Dev, 124*(10-12), 989-998.

Luczak, M. W., & Jagodzinski, P. P. (2006). The role of DNA methylation in cancer development. *Folia Histochem Cytobiol, 44*(3), 143-154.

Matsumura, N., Huang, Z., Mori, S., Baba, T., Fujii, S., Konishi, I., et al. (2011). Epigenetic suppression of the TGF-beta pathway revealed by transcriptome profiling in ovarian cancer. *Genome Res, 21*(1), 74-82.

Michaelson-Cohen, R., Keshet, I., Straussman, R., Hecht, M., Cedar, H., & Beller, U. (2011). Genome-wide de novo methylation in epithelial ovarian cancer. *Int J Gynecol Cancer, 21*(2), 269-279.

Nephew, K. P., & Huang, T. H. (2003). Epigenetic gene silencing in cancer initiation and progression. *Cancer Lett, 190*(2), 125-133.

O'Brien, C. A., Pollett, A., Gallinger, S., & Dick, J. E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature, 445*(7123), 106-110.

Okano, M., Takebayashi, S., Okumura, K., & Li, E. (1999). Assignment of cytosine-5 DNA methyltransferases Dnmt3a and Dnmt3b to mouse chromosome bands 12A2-A3 and 2H1 by in situ hybridization. *Cytogenet Cell Genet, 86*(3-4), 333-334.

Ophorst, O. J., Radosevic, K., Klap, J. M., Sijtsma, J., Gillissen, G., Mintardjo, R., et al. (2007). Increased immunogenicity of recombinant Ad35-based malaria vaccine through formulation with aluminium phosphate adjuvant. *Vaccine, 25*(35), 6501-6510.

Passegue, E., Jamieson, C. H., Ailles, L. E., & Weissman, I. L. (2003). Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci U S A, 100 Suppl 1*, 11842-11849.

Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer, 9*(4), 265-273.

Pommier, Y., Sordet, O., Antony, S., Hayward, R. L., & Kohn, K. W. (2004). Apoptosis defects and chemotherapy resistance: molecular interaction maps and networks. *Oncogene, 23*(16), 2934-2949.

Pradhan, S., & Esteve, P. O. (2003). Mammalian DNA (cytosine-5) methyltransferases and their expression. *Clin Immunol, 109*(1), 6-16.

Prince, M. E., Sivanandan, R., Kaczorowski, A., Wolf, G. T., Kaplan, M. J., Dalerba, P., et al. (2007). Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc Natl Acad Sci U S A, 104*(3), 973-978.

Quintana, E., Shackleton, M., Sabel, M. S., Fullen, D. R., Johnson, T. M., & Morrison, S. J. (2008). Efficient tumour formation by single human melanoma cells. *Nature, 456*(7222), 593-598.

Robertson, K. D. (2005). DNA methylation and human disease. *Nat Rev Genet, 6*(8), 597-610.

Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., et al. (2006). TM4 microarray software suite. *Methods Enzymol, 411*, 134-193.

Sharma, S., Kelly, T. K., & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis, 31*(1), 27-36.

Siddik, Z. H. (2003). Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene, 22*(47), 7265-7279.

Strauss, R., Li, Z. Y., Liu, Y., Beyer, I., Persson, J., Sova, P., et al. (2011). Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PLoS One, 6*(1), e16186.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A, 102*(43), 15545-15550.

Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A, 107*(35), 15449-15454.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., et al. (2010). Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res, 20*(4), 440-446.

Tone, A. A., Begley, H., Sharma, M., Murphy, J., Rosen, B., Brown, T. J., et al. (2008). Gene expression profiles of luteal phase fallopian tube epithelium from BRCA mutation carriers resemble high-grade serous carcinoma. *Clin Cancer Res, 14*(13), 4067-4078.

Vaisman, A., Varchenko, M., Umar, A., Kunkel, T. A., Risinger, J. I., Barrett, J. C., et al. (1998). The role of hMLH1, hMSH3, and hMSH6 defects in cisplatin and oxaliplatin resistance: correlation with replicative bypass of platinum-DNA adducts. *Cancer Res, 58*(16), 3579-3585.

Valcourt, U., Kowanetz, M., Niimi, H., Heldin, C. H., & Moustakas, A. (2005). TGF-beta and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Mol Biol Cell, 16*(4), 1987-2002.

Verdone, L., Caserta, M., & Di Mauro, E. (2005). Role of histone acetylation in the control of gene expression. *Biochem Cell Biol, 83*(3), 344-353.

Wang, S. (2011). Method to detect differentially methylated loci with case-control designs using Illumina arrays. *Genet Epidemiol*.

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., et al. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene, 236*(1), 87-95.

Zhang, S., Balch, C., Chan, M. W., Lai, H. C., Matei, D., Schilder, J. M., et al. (2008). Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer Res, 68*(11), 4311-4320.

Rohit R Jadhav

548 N Senate Ave, # 209, Indianapolis, IN 46204
Phone 201-546-0282, e-Mail **rrjadhav@iupui.edu**

OBJECTIVE

To work and excel in the field of Bioinformatics

EDUCATION

Indiana University, School of Informatics, Indianapolis, IN
Master of Science in Bioinformatics, May 2011, Dean List, GPA: 3.64/4.0
Course Work: Introduction to Bioinformatics, Introduction to Informatics, Biostatistics, Translational Bioinformatics, Biological database management, Informatics Research Design, Algorithms for Bioinformatics, Seminars for Bioinformatics, Machine Learning.

Dr. D. Y. Patil Biotechnology and Bioinformatics Institute Pune, India
Bachelor of Technology in Bioinformatics, June 2009, GPA: 3.16/4.0
Course Work: Molecular Biology, Microbiology, Immunology, Biochemistry Protein, Modeling, Structural Bioinformatics, Chemoinformatics, Molecular Dynamics and Simulations, Molecular Docking, Virtual Screening, High Throughput Screening

Additional Industrial Training:
Design & Production of Bacterial vaccines - Serum Institute of India Pvt. Ltd Pune
C, C++, Java and Advanced Java, Certificate courses

PROJECT EXPERIENCE

Graduate Thesis: Epigenetic analysis of Ovarian Cancer initiating cells

- Analyzing the differences in DNA methylation, gene expression and miRNA patterns in Ovarian Tumor cells, Ovarian Cancer initiating cells and normal cells.
- Integrating the information from above experiments to identify biomarkers.
- Validating the identified biomarkers with the publicly available data.

Implementing Algorithms in Bioinformatics

- Implemented an algorithm for Profile Hidden Markov Model which aligns a given DNA sequence with a profile provided by the user.
- Implemented an algorithm for Greedy Motif Search which searches multiple protein sequences for top 20 motifs using a greedy search algorithm.

Diabetic Proteins Repository

- Built a simple and efficient interface to be used by scientists working on Diabetes.
- Used different repositories like Uniprot, KEGG, Drug Bank for information collection.
- Collected & Integrated information for proteins, pathways & drugs related to Diabetes.
- Created a database and interface for the information representation.

Decision support system for clinicians prescribing antibiotics

- Built an interface to be used by physicians while prescribing antibiotics to patients.
- Collected data for symptoms, diagnosis and antibiotics for different diseases.
- Integrated information into a database with corresponding tables and relationships.
- Created a user interface for simple and efficient use by the physicians.

Undergraduate Thesis: Sequence, Structure and Phylogenetic Analysis of GH Family 10, 11 Xylanases in Fungi

- The comparative study of the Fungal Glycosyl Hydrolase Family 10 and 11 xylanases.
- Identified and selected xylanases unique to the GH Families 10 and 11 from Fungi.
- Compiled and curated the molecular data for the components of the Families 10 and 11
- Performed Sequence and Structure Alignment of the residues.
- Performed molecular Phylogenetic study to get an insight into the evolutionary trend of xylanases belonging to these Families in Fungi.

WORK EXPERIENCE

Indiana University Purdue University Indianapolis

Center for Computational Biology and Bioinformatics (September 2010 - Present)

- Working on analysis of High-Throughput Sequencing, Next Generation Sequencing data.
- Working on analysis of DNA methylation, mRNA & miRNA expression data.

Teaching Assistant (September 2010 - Present)

- Worked as a teaching assistant for graduate course in Introduction to Informatics.
- Conducted lectures on introduction to programming languages and databases.
- Assisted on evaluations of assignments and examinations.

Website Developer, University Information Technology Services
(November 2009 - August 2010)

- Worked with the Podcasting team.
- Worked on development of Administrator interface in Perl, Java, Asp.Net and C#

Internship - University Information Technology Services (Summer 2010)

- Developed interfaces in .NET and Perl-CGI.
- Worked with XML, HTML, JavaScript.
- Worked on implementing client and server side cookies in JavaScript and Perl.

TECHNICAL SKILLS

Programming Languages/Scripting Languages: R, Perl, Python, C, C++, Java, Visual Basic, SQL, DBMS, C#, Asp.Net.

Database Applications: SQL, SQL/PL, Access, Database Design and Architecture

Software Applications/Programs: Partek, Toucan, MOE, Hyperchem, Schrödinger, Autodock 3.5, Pymol

Operating Systems/Platforms: .NET, Computer Fundamentals, Networking, UNIX and Windows 95/ 98/ 2003/ NT/ 7, LINUX, MAC, MS Office - All Word Processes, MS Excel, MS Power point.

LANGUAGES

Fluent: English, Hindi, Marathi. **Conversational**: German

ACTIVITIES

- Volunteer, Blood Donor
- Member, Indian Student Association and Maharashtra Mandal
- Table - Tennis Team member, Dr. D. Y. Patil Biotechnology and Bioinformatics Institute
    - Won singles and doubles tournaments at college and district level in Table Tennis
- Team Member, additional sports (Swimming, Cricket, Badminton etc.)