EXPLORATION OF METHYLATION-DRIVEN MECHANISMS IN CANCER

by

BUĞRA ÖZER

Submitted to the Graduate School of Engineering and Natural Sciences

in partial fulfilment of the requirements for the degree of

Doctorate of Philosophy

SABANCI UNIVERSITY

April 2016

EXPLORATION OF METHYLATION-DRIVEN MECHANISMS IN CANCER

APPROVED BY:

Prof. Dr. Ugur Sezerman
(Dissertation Co-advisor)

Prof.Dr. Ismail Cakmak
(Dissertation Co-advisor)

Prof. Dr. Yucel Saygin

Assoc. Prof. Devrim Gozuacik

Assoc. Prof. Oguzhan Kulekci

DATE OF APPROVAL: 04.04.2016

EXPLORATION OF METHYLATION-DRIVEN MECHANISMS IN CANCER

Bugra Ozer

Molecular Biology, Genetics & Bioengineering, PhD Thesis, 2016

Thesis Advisor: Prof. Dr. Ugur Sezerman & Prof. Dr. Ismail Cakmak

Key words: Transcriptomics, Epigenetics, Data analysis, Data integration, Cancer Biology, Protein-protein interaction network, Functional Enrichment Analysis

ABSTRACT

DNA methylation is an important epigenetic phenomenon that plays a key role in the regulation of expression. For this reason, there have been many studies on the topic of methylation's role in cancer mechanisms. These studies include analyses based on differential methylation, with the integration of expression information as supporting evidence. In the present study, we firstly focused on defining an optimal analysis strategy when both expression and methylation information are available. We investigated the methylation and expression changes on the genes themselves to have a deeper knowledge of thyroid cancer etiology. Moreover, we explored more important genomic regions considering methylation information and identified common and distinct genes and pathways among different cancer types. In addition, we defined a novel graph-based analysis strategy for identifying methylation-driven potential cancer-causing gene patterns. We applied our method to variety of cancers using the Illumina HumanMethylation450k methylation chip and RNA sequencing data. To extract the significantly altered methylation-driven patterns within a STRING protein-protein interaction network, we first defined a methylation change threshold for "large methylation changes". Subsequently, in addition to focusing on the interplay between methylation and expression, we carefully considered the individual relationships between different genes to ensure a deeper understanding of the methylome and transcriptome. Furthermore, we studied the presence of shared and distinct features among the different

types of cancers using hierarchical clustering analysis. Overall, our work not only defined a novel approach for the identification of significantly altered methylation-driven pathways but it also contributed to improving our knowledge of the etiologies of different cancers and the common and distinct features among them.

KANSERDE GÖZÜKEN METİLASYON SEBEPLİ DEĞİŞİMLERİN ARAŞTIRILMASI

Buğra Özer

Moleküler Biyoloji, Genetik ve Biyomühendislik, Doktora Tezi, 2016

Tez Danışmanları: Prof. Dr. Uğur Sezerman & Prof. Dr. İsmail Çakmak

Anahtar kelimeler: Transkriptomiks, Data analizi, Data entegrasyonu, Kanser, Kanser Biyolojisi, Protein-protein etkileşim ağı, Fonksiyonel zenginleştirme analizi

ÖZET

Son yıllarda gelişen teknoloji ile beraber, DNA metilasyonu'nun önemi daha iyi anlaşılmıştır. Özellikle metilasyondaki değişimlerin gen ekspresyon seviyelerini düzenlemesi, ve bu değişimlerin farklı hastalıklarla ilişkilendirilmesi, bu alanda yapılan çalışma sayısını ciddi miktarda yükseltmiştir. Literatürdeki önceki çalışmalara bakıldığında, metilasyon tabanlı çalışmaların genellikle ekspresyon çalışmaları ile beraber yürütüldüğü görülmüş ve metilasyon analizinden çıkan sonuçların ekspresyon sonuçlarıyla karşılaştırılması seklinde bir analiz yapısı oluşturulduğu gözlemlenmiştir. Yaptığımız çalışmada metilasyonun çeşitli kanserler üzerindeki etkilerini, Illumina humanMethylation450k Chip ve RNA sekanslama verileri kullanarak araştırdık. Daha önceki çalışmalardan farklı olarak, öncelikle hem metilasyon hem de ekspresyon bilgilerinin bulunduğu çalışmalarda en uygun analiz nasıl yapılır konusuna eğilerek, yeni bir analiz metodu tanımladık. Daha sonra bu metodu dört farklı kanser türüne uygulayarak, hem kanserler arasındaki ortak ve farklı olarak gözüken değişimleri hem de metilasyon değişimi gözüken genlerdeki daha önemli genomik bölgeleri tanımlamaya calıştık. Bunlara ek olarak, protein-protein etkileşim ağını da analize dahil ederek, sadece genlerin kendilerindeki değişimin değil, çevrelerinde gerçekleşen yüksek metilasyon değişimlerinin de düşünüldüğü yeni bir analiz metodu tanımladık. Sonuç olarak elde ettiğimiz bulgular, kanser biyolojisi ve olası kanser terapi metotları açılarından yüksek önem barındırmaktadır. Gelecekte yapılacak destekleyici çalışmalarla beraber, çıkarttığımız bulguların kanserdeki metilasyon tabanlı değişimleri anlamak açısından açıklayıcı olacağını ön görmekteyiz.

“To my family and close friends”

ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis supervisor Prof. Dr. Ugur Sezerman for supporting me with a great patience throughout this study. His guidance and inspiration have provided and invaluable experience that will help me in my career.

I would like to express my thanks to the thesis committee: Prof. Dr. Ismail Cakmak, Prof. Dr. Devrim Gözüaçık, Prof. Yücel Saygın and Assoc Prof. Dr. Oguzhan Kulekci, for their invaluable review.

I would like to express special thanks to all Sezerman lab members for technical and moral support. I also thank the Professors, fellow graduate students and staff at molecular biology, genetics & bioengineering program.

Last but not the least; I would like to thank my parents Seyhan and Serafettin Ozer, sister Sena Eskiil, and my wife Elif Damla Ozer for their unconditional love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

TCGA – The Cancer Genome Atlas

THCA - Thyroid carcinoma

BRCA – Breast carcinoma

PRAD – Prostate adenocarcinoma

COAD – Colon adenocarcinoma

LUSC – Lung squamous cell carcinoma

CHOL – Cholangiocarcinoma

KIRP – Kidney renal papillary cell carcinoma

LIHC – Liver hepatocellular carcinoma

MDS – Multidimensional scaling plot

FDR – False discovery rate

BH – Benjamini-Hochberg

DNA – Deoxyribonucleic acid

RNA – Ribonucleic acid

RNASeq – RNA sequencing

5-aza-CR - 5-aza-2′-azacytidine

5-aza-CDR - 5-aza-2′-deoxycytidine

HDAC – Histone deacetylase

miRNA – MicroRNA

CpG – 5Cytosine-Phosphate-Guanine3

PPI – Protein protein interaction

MAPK – Mitogen activated protein kinase

5'UTR – Five prime untranslated region

3'UTR – Three prime untranslated region

BMIQ – Beta-mixture quantile

SNP – Single nucleotide polymorphism

CNA – Copy number aberration

RSEM – RNAseq by expectation-maximization

PANOGA – Pathway and network oriented genome wide association study analysis

KEGG – Kyoto encyclopaedia of genome and genes

GO – Gene ontology

EGFR – Epidermal growth factor receptor

TGF – Transforming growth factor

VEGF – Vascular endothelial growth factor

1stExon – First Exon

TSS200 – Near 200 base pairs of transcription start site

TSS1500 – Near 1500 base pairs of transcription start site

BP – Base pair

ECM – Extracellular matrix

GPCR - G protein-coupled receptor

DMR - Differentially methylated regions

DE – Differentially expressed

MVP – Methylation variable position

TSGene – Tumor suppressor gene database

PCR – Polymerase chain reaction

|2| - absolute value of two

> - greater

< - smaller

= - equals

1. INTRODUCTION

1.1 Cancer Background

1.1.1 Cancer Statistics

Cancer is defined as a large group of diseases that can smite any part of the body due to the unregulated cell growth. According to 2013 US statistics, cancers are the second most frequently observed cause of death with 584,551 cases, where there is only 20,000 cases difference with heart diseases, which is the most common cause of death in US (Group, 2014). These numbers are also reflected at worldwide statistics, as there were approximately 14 million newly diagnosed patients and 8.2 million cancer related deaths in 2012, as cancers were one of the leading grounds of morbidity and mortality worldwide. More importantly, the number of new cases is predicted to increase by about 70% over the next 2 decades (McGuire, 2016). Compared to other cancer types, lung cancer is the most common cancer observed worldwide with 1,825,000 newly diagnosed patients (Ferlay et al., 2015) (Figure 1-1 and 1-2). As illustrated at, breast, colorectal and prostate cancers are following lung cancer by the means of new cases.

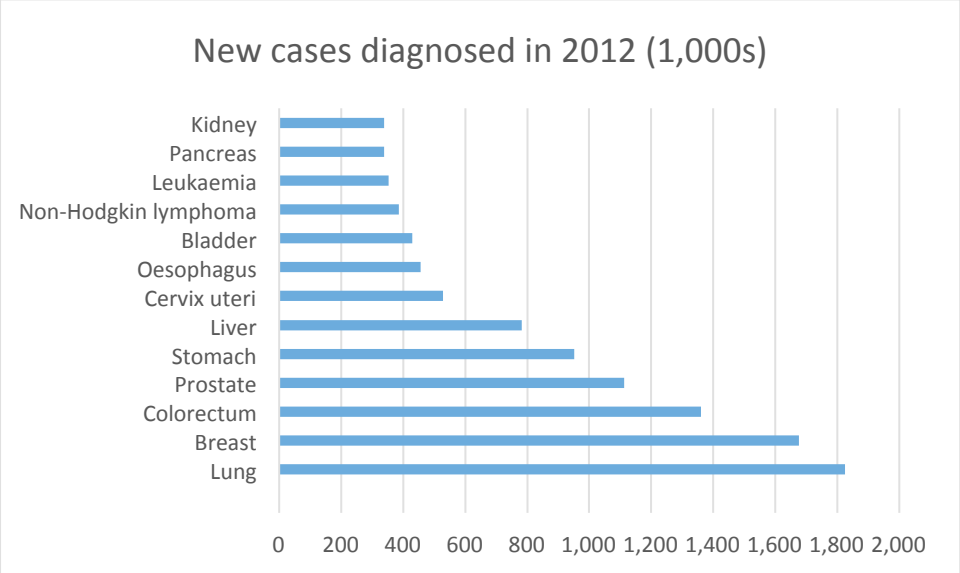

Figure 1-1 Number of new cases diagnosed as cancer in 2012

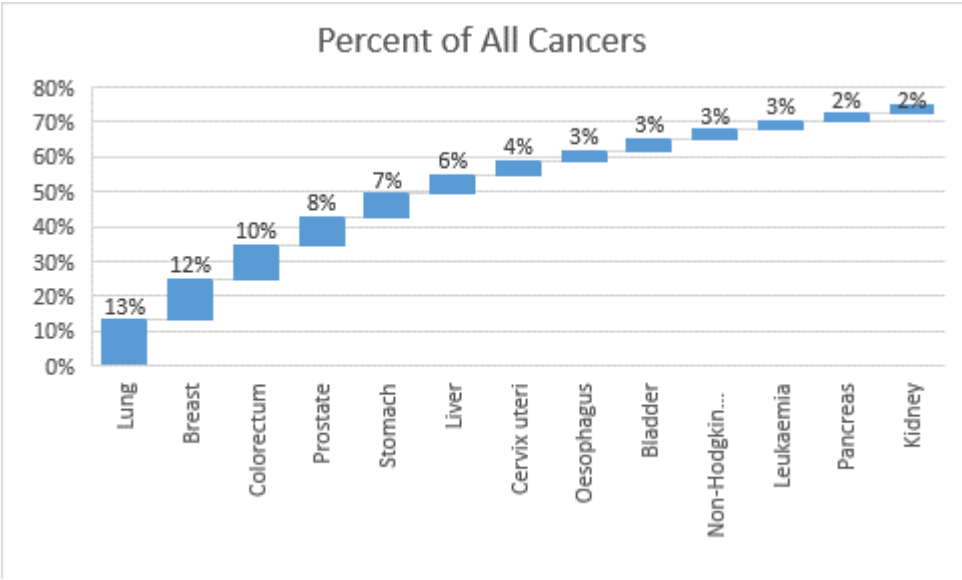

Figure 1-2 Percentage of cancer cases compared to total number of cancer cases

When focused on the mortality rate associated with each cancer type, lung cancer displays highest mortality rate with over 1,500,000 deaths, whereas breast and prostate cancer has lower mortality rate since surgical treatment cures a high percentage of patients diagnosed with these cancer types Figure 1-3.

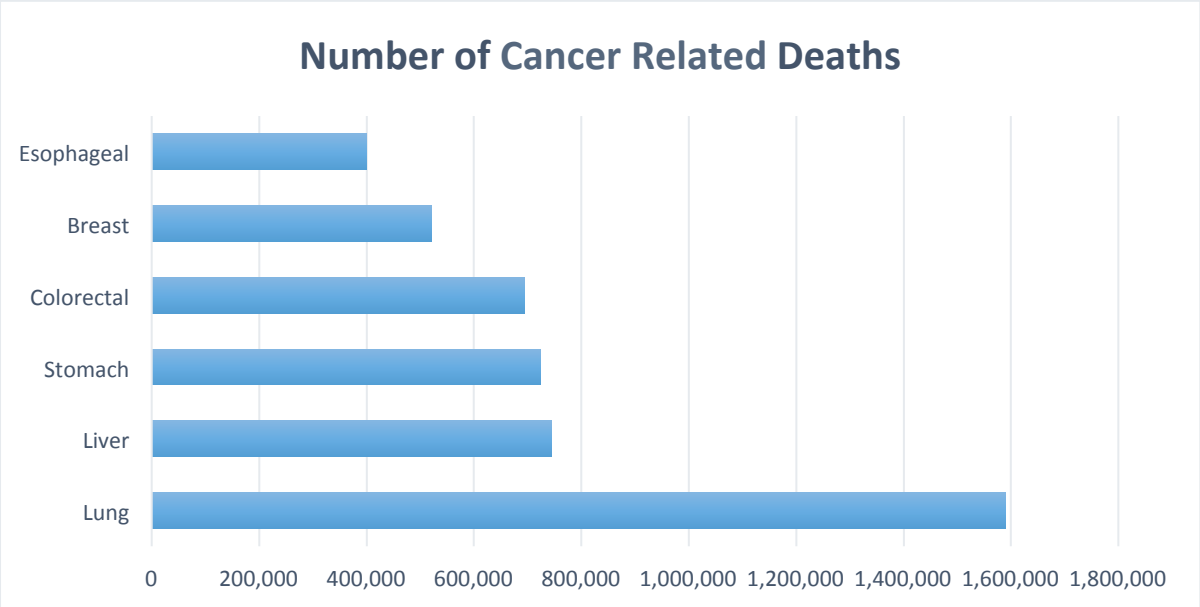

Figure 1-3 Number of cancer related deaths in 2009

In Turkey, according to 2009 statistics 21% of the deaths are caused by cancers, where highest percentage of cases belonged to lung cancer. Similar to the worldwide cancer

distribution, breast, prostate and colorectal cancers were observed frequently in Turkey. Interestingly, bladder cancer was observed at higher rankings compared to worldwide statistics (Yilmaz et al., 2011).

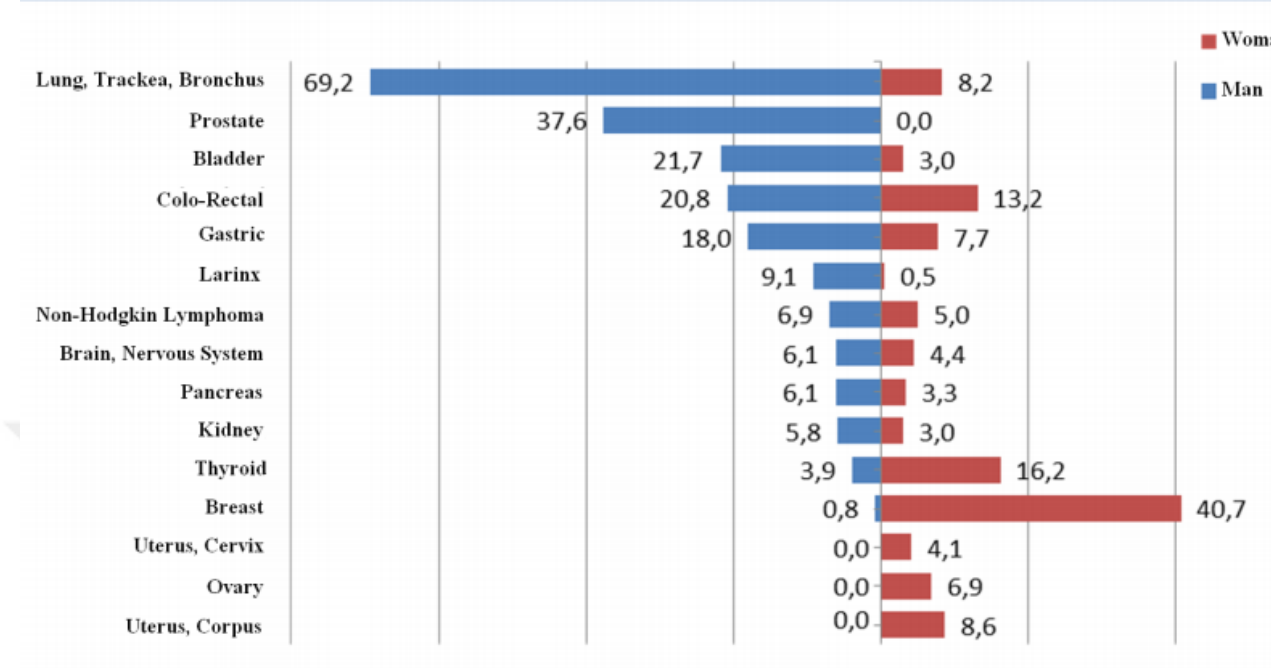

Figure 1-4 Male and female occurrence rates of cancers that are witnessed in 2009 in Turkey.

Moreover, lung, prostate and bladder cancers were observed more frequently in man, whereas thyroid, breast, and uterus, ovary cancers were observed more frequently in woman, according to 2009 Turkey statistics (Figure 1-4).

1.1.2 Cancer Biology

Basically, cancer can be defined as a disease in which set of irregular cells develop irrepressibly by ignoring the standard rules of cell division. Normally, our cells are regularly subject to division, differentiation and death signals. Interestingly, cancer cells develop independence from these signals, resulting in uncontrolled cell growth and proliferation.

Biological causes of cancer cells can be summarized under six topics; immortality of cells (continuous cell division and proliferation), uncontrolled growth signals from oncogenes, overriding stop signals of tumour suppressors, resistance to cell death (apoptosis), metastasis (spread to other tissues) and angiogenesis (induction of new blood vessel growth) (Hejmadi, 2010). Although all cancer cells share these six properties, mechanisms leading to these consequences vary for each cancer type and whether there is a shared set of gene interactions

leading to cancer state remains unknown. In order to improve life standards and to tackle cancer more efficiently, there is a need to understand the detailed cancer mechanisms, which would shed light into the improved cancer therapeutics.

1.2 Epigenetics and cancer

Chromatin structure basically describes how genetic information is organized within a cell, influencing whether genes should get silenced or activated. The current definition of epigenetics involves modifications on chromatin structure and it is defined as 'the study of heritable changes in gene expression that occur independent of changes in the primary DNA sequence' (Sharma, Kelly, & Jones, 2010). In fact, although the genetic information in the DNA is same, epigenetic changes enable cells to have diverse properties.

Genetic factors are well-known to play role at cancer initiation and progression (Jones & Baylin, 2002, 2007). It is now well established that cancer-causing mutations differ from cancer to cancer. Overall, the genetic background of cancer is widely accepted; however, recent studies also suggest that epigenetic alterations may be the vital triggering factors in differing cancer types (Feinberg, Ohlsson, & Henikoff, 2006). Recent innovations in the field of epigenetics have exposed that various epigenetic aberrations are observed at human cancer cells (Sharma, et al., 2010). Basically, these genetic and epigenetic modifications interact at different stages of cancer development, so that cancer progression is promoted as a cumulative effect (Jones & Laird, 1999). Consequently, number of efforts on understanding the role of epigenetics in the initiation and propagation of cancer has rapidly increased (Jones & Martienssen, 2005). Since it has been previously proven that epigenetic aberrations are potentially reversible unlike genetic mutations and the epigenome can be restored to their normal state (C. B. Yoo & Jones, 2006), initiatives on investigating epigenetic grounds of different cancer types is promising and therapeutically significant.

1.3 Effect of DNA methylation on cancer etiology

Most common epigenetic modification observed in our cells is the methylation of cytosine bases in DNA. The general idea about the effect of methylation on expression is that

methylation plays crucial roles in gene silencing and the regulation of gene expression. It is well established that hypermethylation in promoter regions leads to inactivation, whereas hypomethylation is associated with genomic instability and loss of imprinting, which may be the key factors in the reproduction and metastasis of cancer cells (Fraga et al., 2004). Indeed, both hyper- and hypomethylation have previously been associated with a variety of cancers, including kidney, colon, pancreas, liver and lung cancers (Akiyama, Maesawa, Ogasawara, Terashima, & Masuda, 2003; Badal et al., 2003; Cui et al., 2002; Esteller, 2007, 2008; Feinberg & Tycko, 2004; Galm, Herman, & Baylin, 2006; Herman & Baylin, 2003; Herman et al., 1994; Wang, Li, & Wen, 2005; H. Wu et al., 2005).

1.4 Epigenetic therapy of cancer

In cancer, when the causal epigenetic change is detected, there is an option of epigenetic therapy, which enables restoration of healthy epigenome. Recently, many epigenetic drugs have been proposed which can effectively reverse DNA methylation that occur in cancer (C. B. Yoo & Jones, 2006).

When looked at the roots of epigenetic drugs, first ones were the DNA methylation inhibitors, which were designed to reverse the high methylation in cancer cells. Remarkably, in 1977 treatment with cytotoxic agents namely; 5-azacytidine (5-aza-CR) and 5-aza-2′-deoxycytidine (5-aza-CdR), were proposed for inhibition of DNA methylation which would encourage gene expression and trigger differentiation in cultured cells (Constantinides, Jones, & Gevers, 1977). Hence, this was the first evidence for potential use of epigenetic drugs in cancer therapy. Especially drug-induced reduction of DNA methylation leads to activation of previously silenced tumor suppressors and growth of cancer cells is inhibited by that way leading to potential therapy. Biological-wise, basically these DNA methylation inhibitor drugs show their affect by trapping DNA methyltransferases onto the DNA, thus blocking the methylation procedure.

However, one must also consider drug toxicity while designing drugs. From this perspective, methylation-reversing drugs mainly target rapidly dividing tumor cells and have minimal effects on slowly dividing normal cells since they interfere the cell at replication stage. This argument has been supported by various studies demonstrating minimal side

effects of long-term treatment with DNA methylation inhibitors (A. S. Yang, Estecio, Garcia-Manero, Kantarjian, & Issa, 2003).

Additionally, it has also been shown that the use of methylation-reversing drugs may be beneficial for effective combinatorial cancer treatment. There have been many previous approaches for this purpose, such as synergistic use of HDAC inhibitors and DNA methylation together. Actually, compared to individual treatment approaches, combination treatment strategies are shown to be more efficient. For example, combinatorial effect of 5-Aza-CdR and trichostatin A are shown to activate putative tumor suppressor genes (Cameron, Bachman, Myohanen, Herman, & Baylin, 1999). Harmonious activities between HDAC inhibitors and DNA methylation were also revealed in a mice study, as treatment of phenylbutyrate and 5-Aza-CdR led to larger reduction of lung tumor formation (Belinsky et al., 2003).

The use of miRNAs for epigenetic therapy is also validated by various studies. The finding by Saito *et al.* (Saito et al., 2006) proved that after treatment with 4-phenylbutyric acid and 5-Aza-CdR, reactivation of *miR-127* inhibits BCL6 oncogene. In addition, synthetic, mimic tumor suppressor miRNAs are shown to be promising candidates for selectively repressing oncogenes in tumors. Additionally, *miR-101* that targets EZH2 (Friedman et al., 2009) was shown to be useful for regulating the abnormal epigenetic mechanism in cancer. However, the major obstacle of this strategy is the lack of efficient delivery methods, hence development of efficient vehicle molecules for targeted delivery of synthetic miRNAs to tumor cells will be of highest importance in future.

1.5 Methylation – expression relationship & previous works

Previously in the literature, a highly preferred way to focus on methylation-based changes was to conduct differential methylation analysis. Since methylation change shows it effects on transcriptomics level, focusing only on differentially methylated genes is questionable while explaining the disease status. There is a need to incorporate transcriptomics information into the methylation studies to observe what actually happened in the genomics level and this was adopted by many scientists as incorporating both epigenetics and transcriptomics information into disease identification studies would shed light to the disease

etiology, improving the treatment procedure. However, integrating methylation and expression data is a problem that is commonly confronted due to the complex relationship between methylation and expression as a change in methylation level does not always lead to a corresponding change in expression level due to variety of factors.

Although the opposite pattern has also been observed in several studies (Wan et al., 2015; S. Yoo et al., 2015), an inverse correlation is expected between changes in methylation levels and expression (Hovestadt et al., 2014; Smith et al., 2014; van Eijk et al., 2012). A recent study by Lee et al. demonstrated that there is a tendency toward direct correlations in backbone regions, whereas inverse correlations are expected near CpG sites in promoter regions (Lee & Wiemels, 2016). Additionally, gene silencing via the hypermethylation of tumor-suppressing genes and activation of tumor-promoting genes via hypomethylation has been demonstrated to favor oncogenesis (Ehrlich, 2002).

Recent studies show that searching for correlation between methylation and expression data is the most adopted strategy on tackling this problem. In this type of analysis, statistical analysis of both methylation and expression data are conducted separately and at the final stage, these results are compared with each other (Alashwal, Dosunmu, & Zawia, 2012; Fan & Zhang, 2009; Gervin et al., 2012; M. Li et al., 2009; Paziewska et al., 2014; Taskesen et al., 2014). At the work of Alashwal et al. an integrative analysis of global gene expression and methylation profiles is conducted for the purpose of investigating the role of DNA methylation in gene expression regulation in Alzheimer data. At this work, although it has been claimed that methylation and expression data are combined, only combination of data was the comparison of transcriptomics and methylation analysis results. Gervin et al. investigated DNA Methylation and gene expression changes in monozygotic twins discordant for psoriasis. At the study, no gene was identified as differentially methylated or differentially expressed gene. However, when they combined these two data by using Spearman's rho as a measure for correlation, all genes were ranked according to the significance of the correlation coefficients. After that, they were able to link some of the genes with psoriasis. Another example can be the work of Li et al., which involves integrated analysis of DNA methylation and expression in ovarian cancer. Again at this work, there is not any combination of methylation and expression information, instead methylation and expression analysis are handled separately, while these information is combined during biological interpretation stage.

Another approach is to merge methylation and expression data prior to any kind of analysis by implementing general data integration algorithms. For this purpose, Kim et al. benefited from graph integration using Laplacians to merge multi-layer genomic data in glioblastoma clinical outcome prediction. The formula they have used for this purpose can be observed below:

$$\min_{\alpha} y^T \left(I + \sum_{k=1}^{K} \alpha_k L_k\right)^{-1} y, \quad \sum_k \alpha_k \leqslant \mu$$

where K is the number of graphs (data sources) and L_k is the corresponding Laplacian of graph G_k. Moreover, last solution is obtained for repeating the first formula for each layer and using the below formula.

$$f = \left(I + \sum_{k=1}^{K} \alpha_k L_k\right)^{-1} y$$

On the other hand, Taskesen et al. benefited from clustering of separate methylation and expression analyses. Although these computation-based methods may be beneficial to reveal new insights, since they neglect the nature of biological relationship between methylation and expression, they do not yield optimal results, hence there is a need to define the optimal analysis strategy when both methylation and expression data are available.

Efforts have also been focused on identifying the pathways that are mainly affected by methylation-driven changes. In a recent study, Gevaert et al. developed a univariate beta mixture model-based method for the identification of differential methylation, termed MethylMix (Gevaert, Tibshirani, & Plevritis, 2015), to explore transcriptionally predictive methylation-driven genes and pathways in twelve different cancers. Basically, MethylMix involves three-step algorithm: Firstly, genes are annotated as “transcriptionally predictive”, depending on methylation of CpG sites effect on expression. Secondly, methylation states of a gene are identified using univariate beta mixture modeling to check whether subgroups of patients share similar DNA methylation levels. Thirdly, Wilcoxon rank sum test is applied to define hyper and hypomethylated genes by comparing tumor tissue results with normal tissue results. After then, differential methylation values are assigned using the difference of methylation state with the normal methylation state. Genes that are significantly different are identified as “differential genes”.

Alternatively, Kim et al. proposed a logistic regression-based method for gene set enrichment, termed LRpath (Kim et al., 2012), for the investigation of important methylation-driven pathways. Moreover, pathway similarities across different types of cancers have also been included in this analysis.

1.6 Research Aim and Goal

To study the role of methylation in cancer development mechanisms, one must explore how oncogenes and tumor suppressor genes (drivers) are modified or how differentially methylated genes alter the expression levels of driver genes through a set of interactions in a protein-protein interaction (PPI) network. In the present study, we firstly focused on defining an optimal analysis strategy when both expression and methylation information are available while identifying the methylation and expression changes on the genes themselves to have a deeper knowledge of cancer etiology. Moreover, we investigated more important genomic regions considering methylation information and identified common and distinct genes and pathways among different cancer types. In addition, we defined a novel graph-based analysis strategy for identifying methylation-driven potential cancer-causing gene patterns. We applied our method to variety of cancers using the Illumina HumanMethylation450k methylation chip and RNA sequencing data. To extract the significantly altered methylation-driven patterns within a STRING protein-protein interaction network, we first defined a methylation change threshold for “large methylation changes”. Subsequently, in addition to focusing on the interplay between methylation and expression, we carefully considered the individual relationships between different genes to ensure a deeper understanding of the methylome and transcriptome. Furthermore, we investigated the presence of shared and distinct features among the different types of cancers using hierarchical clustering analysis. Overall, our study not only defined a novel approach for the identification of significantly altered methylation-driven pathways but also contributed to improving our knowledge of the etiologies of different cancers and the common and distinct features among them.

2. METHODS

2.1 Dataset

Dataset consisting of 8 normal and 10 tumour samples are obtained from Batch230 and dataset consisting of 6 normal and 6 tumour samples are obtained from Batch250 Thyroid Cancer Carcinoma data in The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research, 2014). This dataset was used as a case study and training dataset. Additionally, we have also downloaded another 30 samples from the same source to test our findings on another independent dataset.

Additionally we have used another dataset consisting of 128 normal and 128 tumour samples from 4 different cancer types namely; breast carcinoma (BRCA, 84 samples) (Cancer Genome Atlas, 2012b), thyroid carcinoma (THCA, 60 samples) (Cancer Genome Atlas Research, 2014), colon adenocarcinoma (COAD, 52 samples) (Cancer Genome Atlas, 2012a) and prostate adenocarcinoma (PRAD, 60 samples) (Taylor et al., 2010).

Lastly, we conducted analyses of methylation and expression data from six different cancer, which included thyroid cancer (Cancer Genome Atlas Research, 2014), lung squamous cell carcinoma (Cancer Genome Atlas Research, 2012), kidney renal papillary cell carcinoma (Cancer Genome Atlas Research et al., 2016), colon adenocarcinoma (Cancer Genome Atlas, 2012a), cholangiocarcinoma (provisional), and liver hepatocellular carcinoma (provisional). While treating each cancer type individually, we selected only samples that were matched to the anatomic site of the tumor. Moreover, because our approach was sensitive to possible noise factors due to the integration of methylation and expression information, we only included samples with available data on both control and tumor samples. We included a total of 92 THCA samples, 18 CHOL samples, 30 COAD samples, 46 KIRP samples, 14 LUSC samples and 78 LIHC samples in the datasets. The methylation data consisted of intensity values matching CpG sites covering different regions of the gene, whereas the RNA-Seq data consisted of count values corresponding to each gene that were computed by the data owners.

We have only continued our analysis with the samples that contain both RNA sequencing and methylation data. According to data providers, all methylation data was obtained from Illumina Human Methylation 450k Chip, whereas all RNA sequencing data was obtained from Illumina HiSeq machine. Data consisting of intensity values corresponding to each

region for methylation chip and counting values corresponding to each gene for RNA-Seq are downloaded for our study.

For both methylation and RNA sequencing (RNA-Seq) experiments, statistical analyses are conducted for each batch independently and also by pooling both batches together before pre-processing the data.

2.2 Methylation Analysis

Methylation is a region-specific, rather than a gene-specific phenomenon; hence, methylation in different gene regions can lead to diverse consequences. In our methylation analysis, we benefited from the ChAMP pipeline (Morris et al., 2014), which included in the R-Bioconductor package and is specifically designed for the analysis of Illumina HumanMethylation450k chip data. ChAMP employs a sliding window approach (Probe Lasso) for annotating CpG regions with genomic locations (Butcher & Beck, 2015). CHAMP allows investigation of methylation occurring in different genomic regions, including in the first exon, 3'UTR, 5'UTR, gene body, intergenic region and within 200 bp and 1500 bp proximities of the transcription start sites. Moreover, the beta values associated with each methylation change are used as estimates of methylation levels, which is the ratio of the methylation probe intensity to the overall intensity and provides an intuitive biological interpretation.

In array-based methylation experiments, both Beta-value and M-value statistics are used as metrics to measure methylation levels. Beta-Value in methylation experiments is the estimate of methylation level using the ratio of the methylation probe intensity and the overall intensity whereas M-value is a logit transformation of Beta-Value. For easier functional interpretation of the results, we have used Beta-Value at our analysis, which provides more intuitive biological interpretation as it roughly corresponds to the percentage of a methylation on a specific site (Du et al., 2010).

After downloading the methylation intensity data from TCGA, the BMIQ normalization method (Teschendorff et al., 2013) was applied to avoid the bias introduced by the Infinium type 2 probe design.

After BMIQ normalization, the magnitude of the batch effects was corrected using the ComBat normalization method, which is an empirical Bayes-based method of correcting for

technical variation related to a slide (Johnson, Li, & Rabinovic, 2007). Moreover, because possible polymorphisms in an individual's genome may affect the methylation status of probes, we excluded SNPs with frequencies greater than 0.05 based on the 1000 Genomes Project (Genomes Project et al., 2015).

To avoid false-positive results in the differential methylation analyses, the Benjamini-Hochberg calculation (Benjamini Y, 1995) was applied for all p-values.

Additionally, after pre-processing, analyses of copy number aberrations (CNA) and the segmentation of methylation variable positions (MVPs) into biologically relevant differentially methylated regions (DMRs) were conducted using the "champ.MVP" function of the CHAMP package. When conducting the analysis of copy number aberrations, we focused on the entire gene, rather than only including particular genomic regions. Individual tumor samples were evaluated against pooled normal samples, and the corresponding regions and segmental mean changes were reported in the output of the analysis. To determine whether copy number aberrations led to corresponding expression changes for the same gene, we used a segmental mean change of 2 as the threshold. Moreover, we annotated genes that exhibited both increases and decreases in different samples from the same dataset as "not important".

Moreover, the Illumina HumanMethylation450k chip provides information about more than 450,000 different regions predicted in approximately 22,000 genes in the human body. Consequently, there is usually more than one differentially methylated region that falls within the borders of a given gene, which causes discrepancies in the data. To solve this problem, we evaluated regions of differing methylation within each gene and checked whether the general trend of the change was upregulation or downregulation. Depending on the direction of the change, the CpG region exhibiting the greatest methylation change, and a change in the same direction as the general trend was taken as representing the change in the methylation level for the whole gene.

To investigate the effects of large methylation changes, we first defined "large methylation change" threshold. For this purpose, we pooled all of the data and identified the distribution that best fit the data using the "fitdistr" function of the MASS R package (Marie Laure Delignette-Muller, 2015). Thus, the Cauchy distribution was found to best explain the data. We calculated the central value and scaling parameter for the pooled data as the Cauchy distribution parameters. For central value estimation, we took the truncated mean of the

middle 24% of the sample order statistics, which has been demonstrated to be valid for the Cauchy distribution (Bloch, 1966). In our pooled data, we detected a central value of 16.8%. In contrast, the log-likelihood function was used for the scaling parameter. The corresponding log-likelihood formula can be found below:

$$\sum_{i=1}^{n} \frac{\gamma^2}{\gamma^2 + [x_i - x_0]^2} - \frac{n}{2} = 0$$

where n is the sample size; y is the scaling parameter; and x_0 is the central value. After the calculation of each sample value, we identified a scaling parameter of 7.77. We used the central value plus two scaling parameters away from the center as the "large methylation change threshold"; thus, we set 32.2% as the high methylation threshold.

2.3 RNA sequencing analysis

RNA sequencing analyses for both batches were performed using the edgeR (Robinson, McCarthy, & Smyth, 2010) Bioconductor (Gentleman et al., 2004) package. The raw RNA sequencing reads associated with each sample were not available on the TCGA Server; hence, the quality control, pre-processing, mapping and counting procedures were performed by the providers of the data (Cancer Genome Atlas, 2012a, 2012b; Cancer Genome Atlas Research, 2014; Taylor, et al., 2010). We worked on counting the data produced via the RSEM procedure (B. Li & Dewey, 2011), and we applied EdgeR for the detection of differential expression between the tumor and control samples. EdgeR benefits from empirical Bayes estimation and tests based on the negative binomial distribution (Robinson, et al., 2010). Similar to the methylation analysis, we performed Benjamini-Hochberg corrections (Benjamini Y, 1995) for all p-values.

Additionally, unlike methylation analysis there was not any batch effect problem at our expression analysis, as all downloaded samples were from IlluminaHiSeq_RNASeqV2 of TCGA dataset.

2.4 Combining significance values

For each gene, expression and methylation significances (Benjamini-Hochberg false discovery rates) are combined using survcomp package (Schroder, Culhane, Quackenbush, & Haibe-Kains, 2011) which is a R (R Development Core Team, 2014) package that provides functions to assess and to compare the performance of risk prediction models. In more detail, Fisher's weighted Z-method is applied while merging expression and methylation data.

$$Z_w = \frac{w_X \frac{\sqrt{n_X}\bar{X}}{\hat{S}_X} + w_Y \frac{\sqrt{n_Y}\bar{Y}}{\hat{S}_Y}}{\sqrt{w_X^2 + w_Y^2}}$$

As suggested by Zaykin et al.(Zaykin, 2011) weights are assigned as square root N, where N = sample sizes. Functional enrichment results obtained from separate batches are integrated in a similar fashion.

2.5 Functional Enrichment

Functional analysis for each data set is conducted via PANOGA Functional Enrichment tool (Bakir-Gungor, Egemen, & Sezerman, 2014). PANOGA incorporates protein-protein interaction information while extracting significant pathways. It helps to identify disease related genes and devise functionally essential KEGG pathways through the identification of genes within the pathways.

PANOGA analysis for results of ten different analysis models were conducted using Cytoscape (Shannon et al., 2003). At Cytoscape, we have benefitted from JactiveModules package (Ideker, Ozier, Schwikowski, & Siegel, 2002) and while using JactiveModules "Number of Modules" was set as 1000 and overlap threshold was set as 0.5.

At our analysis, in order to understand the biological distribution of our genes, Gene Ontology (GO) (Ashburner et al., 2000) analysis is conducted using ConsensusPathDB functional annotation tool (Kamburov, Stelzl, Lehrach, & Herwig, 2013). We have used the option of “over-representation analysis” and queried our gene list against Gene Ontology Level 4: Biological Process database with the p-value cut-off of 0.01. Moreover, we have conducted KEGG functional analysis for each term to understand the association between the genes inside of GO terms and the cancer state.

Particularly for our case, significant alteration at post-translational modification and regulation of transcription pathways were of higher importance, as they possess the potential of affecting many biological processes. In order to find out the genes with critical effects, we have searched for transcription factors in TFCat database (Fulton et al., 2009). In addition, functions of the genes that are found as differentially altered and shared by more than 2 cancer types are investigated in more detail using GeneCards (Safran et al., 2010).

In order to test significance of overlap between four different cancer types, we have compared our findings with the probability of observing the same outcome by chance. For this purpose, we have set number of differentially altered genes belonging to each cancer dataset as number of genes to pick randomly from a dataset and repeated this procedure for 10000 times by focusing on the overlap between different cancers. Number of cases with an overlap between four cancer types is reported in results section supporting our findings. As last step, we have searched for commonly affected genes and pathways among different types of cancers using our custom script and Venny software (Oliveros, 2007-2015).

2.6 Analysis performance measure for thyroid cancer data

In order to evaluate different analysis strategies, we have extracted the list of thyroid cancer-related pathways and genes from previous thyroid cancer researches. For each data and significance merging strategy, our main performance measure was to observe thyroid related pathways in top 20 rankings.

On the other hand, for the purpose of understanding whether combining expression and methylation information results in better significance values for thyroid-cancer associated genes, we have compared significances of differential expression, differential methylation and combination of expression and methylation for Batch230, Batch250 and the Pooled dataset. At these tables, we have also included the information of methylation level change for all cases, which is taken as difference in Beta-value corresponding to each gene between two experiment conditions.

2.7 Methylation change threshold analysis

With the aim of comparing the effects of putting various threshold levels for methylation change, a custom script (will be made available upon request) was written which computes

inverse correlation between expression and methylation for all regions in the dataset. Simply, if methylation of a certain gene is upregulated and expression of the same gene is downregulated, that gene is counted as "inversely correlated". Same is applied for vice versa and as a next step, the ratio between number of inversely correlated genes and total number of genes are calculated for all datasets; in our case Batch230, Batch250 and the pooled dataset.

Lastly, the difference in ratio between above and below varying thresholds in the range of 0.05 (5%) and 0.5 (50%) are computed in order to find out the optimal threshold which favours inverse correlation. This way, for example if the threshold is set at 25%; number of regions with methylation change larger than 25% and number of regions with methylation change less than 25% are compared considering the inverse correlation between expression and methylation at that certain gene. Inverse correlation gains corresponding to each dataset is linearly added together and the threshold with highest overall inverse correlation gain was picked as the best-performer.

2.8 Identifying important gene regions

In order to understand the interplay between differential methylation and expression, we have calculated the inverse correlation ratios between methylation change and expression change for six different gene regions (1stExon, 5UTR, 3UTR, TSS200, TSS1500, Gene Body and intergenic region) that are annotated by ChAMP package. We have only focused on regions showing strong inverse correlation with expression. For all cases, average inverse correlation percentages corresponding to each dataset is linearly added together and the threshold with highest overall inverse correlation ratio was picked as the best-performer.

2.9 Extracting suppressors and oncogenes list

In our analyses, to investigate the potentially cancer-causing set of interactions, we searched for validated oncogenes and tumor suppressors in the literature. For this purpose, we benefited from 398 genes that were included in KEGG: Pathways in Cancer list (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016). Moreover, we added 5 genes to this list that were included in the QIAGEN Human Oncogenes and Tumor Suppressor Genes RT^2 Profiler

PCR arrays (QIAGEN, 2016). Overall, among the 403 available genes in the Pathways in Cancer dataset, 129 genes were annotated as either "tumor-suppressors" or "oncogenes" based on the Tumor Suppressor Gene Database (TSGene) (Zhao, Kim, Mitra, Zhao, & Zhao, 2016) and the work of Vogelstein et al. (Vogelstein et al., 2013). Finally, further filtering was conducted because only 110 genes had available methylation values for all cancer types; thus, we used these 110 genes in the correlation analyses and the calculations of average distances from drivers.

2.10 Integrating Protein-Protein Interaction Information

In this study, we extracted methylation-affected cancer-related patterns by examining protein-protein interactions close to the cancer-related genes. For this purpose, we used the widely employed STRING database (Szklarczyk et al., 2015) because STRING covers approximately 9,000,000 proteins and provides information about the types of relationships that exist between pairs of proteins. As the type of interaction was crucial for our analysis, among the eight different interaction types that exist in the STRING database (i.e., activation, binding, expression, post-translational modification, inhibition, phenotype, catalysis and reaction), we considered only "activation" and "inhibition" relationships. Moreover, we filtered out non-human interactions and interactions with low and medium confidence (combined confidence scores<800). Ultimately, a total of 70,518 protein-protein interactions were included in further analyses.

2.11 Driver Distance Calculation

In this work, we focused on identifying methylation-affected patterns that could potentially result in a cancer state. During this process, we utilized a graph-based, multistep approach, as follows:

- First, only genes with expression fold-changes greater than 2 and methylation significances lower than 0.1 were included.
- For each cancer type, we annotated all of the tumor suppressors that were downregulated for that individual cancer type and all of the oncogenes that were upregulated in the same dataset as the "driver genes".

- In the next step, we calculated all of the paths from each gene to these so-called "driver genes" (either oncogenes or tumor suppressors)
- Most importantly, treating these "driver" genes as starting nodes, we traversed all of the interactions from these genes to their 7^{th} neighbors using breadth-first searches. While investigating each level of neighbor, we searched for inverse correlations between expression and methylation.
- Additionally, we considered the "activation" and "inhibition" relationships between interacting genes; for example, if there was an "activation" relationship between two genes, we searched for a direct relationship between the expression levels of the interacting genes, and if an "inhibition" relationship was observed, we searched for an indirect relationship.
- Previously, Ozer et al. demonstrated that setting a 15% methylation change threshold for methylation analysis considerably improves the outcome of the analysis (Ozer & Sezerman, 2015). Thus, to determine whether methylation was the primary reason underlying the observed change in the expression level, we used a 15% methylation change threshold.
- In summary, we utilized three constraints to indicate a path as being "methylation driven":
    - a greater than 15% methylation change
    - an inverse correlation between expression and methylation
    - the "activation" and "inhibition" relationships were preserved for all genes in the pattern
- Importantly, when more than one path was driven by methylation at the same level, that path was also included at our analysis.
- For each driver gene, the methylation-driven paths with the shortest distances were considered, and all paths with longer distances were discarded.
- Finally, to compare high and normal methylation changes, we repeated the abovementioned procedure twice using separate methylation threshold levels: 32.2% and 15%.

To illustrate our proposed method, an example scenario is provided in Figure 2-1.

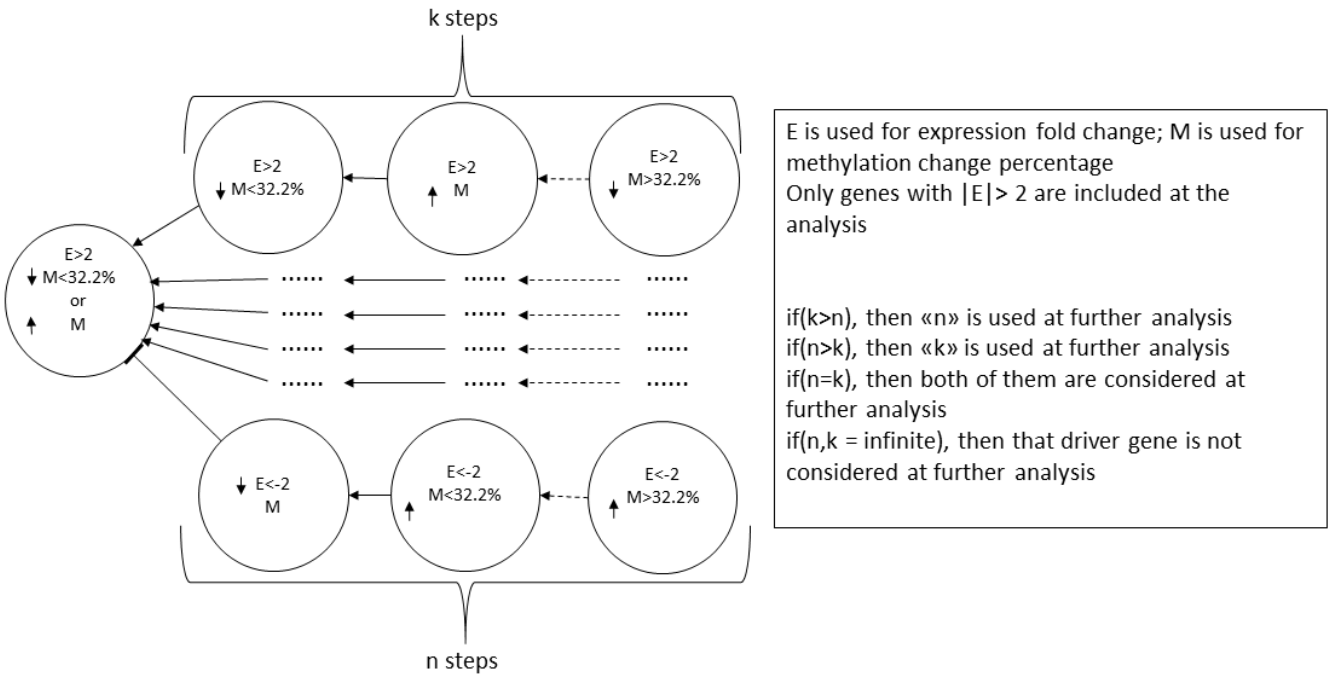

Figure 2-1 Procedure for calculating the number of methylation-driven interaction steps necessary to reach cancer-related genes. This figure was created for oncogenes, and the opposite procedure was applied for tumor suppressors. In this procedure, if the fold-change in an oncogene's expression was >2 or that of a tumor suppressor was <-2, then that gene was added to the short list of cancer driver genes. For each gene on the short list, we searched for a path until we reached a gene showing a change in methylation of >32.2% that caused a corresponding expression fold-change >|2| to calculate the distance for the 32.2% threshold. All of the intermediate steps were required to show cancer-promoting expression changes >|2|.

2.12 Analyzing the similarities of methylation-driven pathways in different cancer types

To determine whether the different types of cancer shared similar methylation-driven gene sets, we defined a measure to assess similarity by focusing on the statuses of the "driver genes" between each cancer. The general scoring scheme for this analysis was as follows:

- If a driver gene itself exhibited a methylation change greater than 32.2% (high methylation) and the direction of this change was inversely correlated with the expression change (log 2 fold-change >1), then that gene was assigned a score of "3". Additional patterns that originated from that gene were not included in the further analysis.
- If a driver gene did not exhibit high methylation or there was no inverse correlation with the expression change, then the neighbors of the driver genes were investigated.

Moreover, if the desired relationship was found, the gene was assigned a score of "2". Additional patterns that originated from that gene and exhibited greater distances from the driver gene were not included in the further analysis.

- If no high methylation-driven pattern was detected at the maximum distance of 2, then the presence of the desired patterns between distances of 3 and 7 was investigated. If any pattern satisfied our rules, -1 was assigned as a penalty, such that more directly interacting genes were favored.
- If no high methylation-driven patterns were found at any distance, then a score of -3 was assigned.

This process was repeated for each cancer type, and after this step, the pairwise differences for each cancer type were calculated and visualized using the R Cluster Package. Additionally, we investigated common cancer mechanisms using only expression or methylation information. For this purpose, we employed only the genes showing non-zero values for all cancer types; a total of 12,109 genes met this criterion. We calculated the pairwise Pearson correlations between each cancer type, and the results were visualized via hierarchal clustering using the "average" distance method. Moreover, to understand the significance of our results, we randomly extracted 110 genes from our dataset and calculated the pairwise correlations 1000 times. We then averaged the correlation results for the 1000 executions to obtain another cluster for comparison with our original findings.

Additionally, the number of shared genes among different cancer types was made available for visualization using the Upset R package (Gehlenborg, 2016).

3. DEFINING OPTIMAL ANALYSIS STRATEGY WHEN BOTH EXPRESSION AND METHYLATION INFORMATION ARE AVAILABLE

3.1 Results

At this analysis, we have only continued with the genes having detected differential methylation and differential expression Benjamini-Hochberg false discovery rates (FDR) below 0.01 to avoid false positive results.

3.1.1 Methylation analysis

While exploring differentially methylated regions, we have only included the regions having a False Discovery Rate (FDR) lesser than 0.01. As a result, we had a list of 1807 significant differentially methylated regions in 1310 different genes for Batch230 and 946 differentially methylated regions in 730 different genes for Batch250. When both batches were pooled together, we were able to obtain 9333 differentially methylated regions in 4729 different genes.

3.1.2 RNA sequencing analysis

Likewise, in RNA Sequencing analysis we have only included genes with False Discovery Rate (FDR) lesser than 0.01. As a result of the analysis, there were 2610 differentially expressed genes for Batch230 and there were 1482 differentially expressed genes for Batch250. When normalized expression values of Batch230 and Batch250 were pooled together (Pooled dataset), we were able to obtain 4790 differentially expressed genes.

3.1.3 Methylation threshold analysis

In order to test whether setting a valid threshold for methylation change yields enhanced functional enrichment results, an analysis is done for varying thresholds between 0.05 (5%) and 0.5 (50%) focusing on inverse correlation between expression and methylation. In both Batch230 and Batch250, genes with methylation change larger than 35% yielded highest ratio (69.23%, 66.45% respectively) of inverse correlation with expression. In the pooled dataset on the other hand, setting 40% methylation change threshold enabled us to reach highest inverse correlation ratio (69.00%) (Figure 3-1).

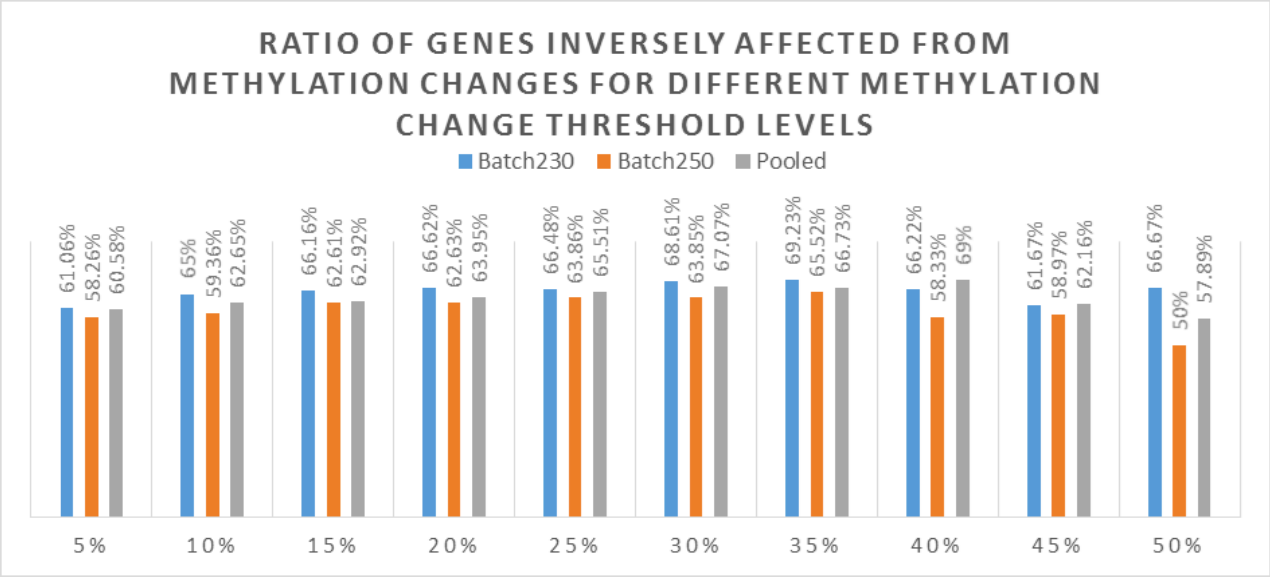

Figure 3-1 Figure showing ratio of inversely affected genomic regions to total number of affected genomic regions with differential methylation (FDR<0.01) for different methylation change threshold values. Results contain analysis for Batch230, Batch250 and pooled dataset separately.

Optimal threshold would be the one that maximizes the difference between ratios above and below of a certain threshold. Although genes having more than 40% methylation change may be informative about the disease state, setting a 40% threshold may not be beneficial for finding the optimal results. At our analysis, highest gain of inverse correlation ratio (29.77%) was obtained by using the threshold of 0.15 (%15) (Figure 3-2).

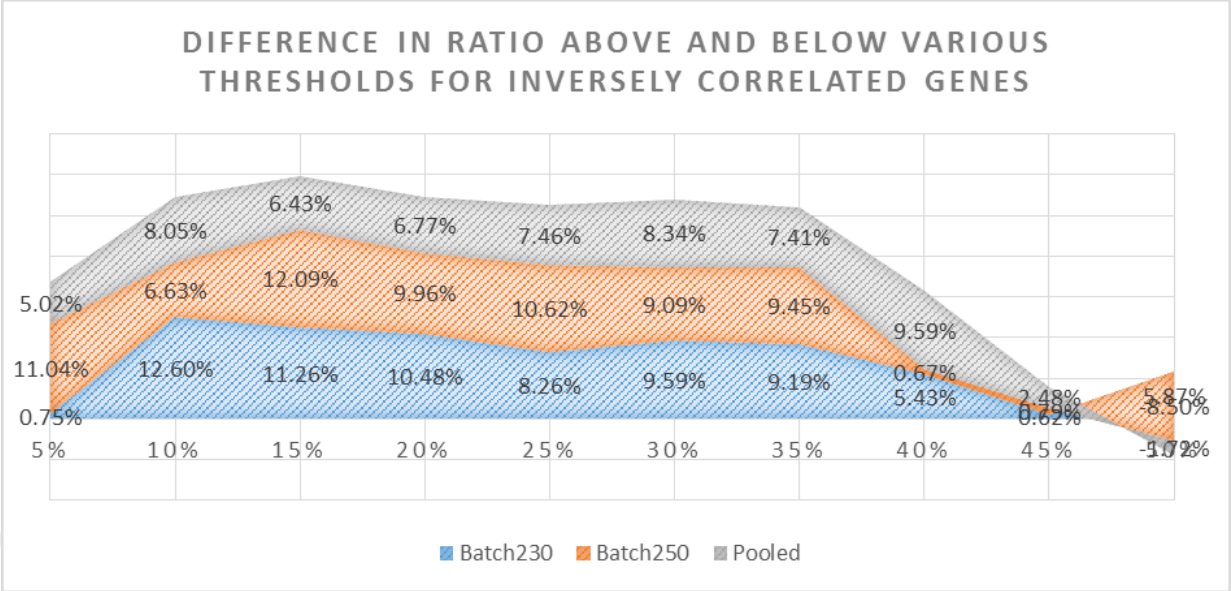

Figure 3-2 Figure showing the difference in ratio of inversely correlated genomic regions (methylation ↑ expression ↓ and vice versa) above and below varying thresholds. Only regions having differential methylation FDR<0.01 are included. Results contain analysis for Batch230, Batch250 and pooled dataset separately.

3.1.4 Functional enrichment analysis

Ten different analysis models and their short summaries are shown in Table 3-1.

Analysis Models	Model Descriptions (FDR<0.01 for all models)
Model 1	Only differentially expressed genes
Model 2	Only differentially methylated regions
Model 3	Differentially expressed and differentially methylated genes
Model 4	Significant genes when methylation and expression significances combined
Model 5	Genes with more than 15% methylation change and inversely correlated with expression
Model 6	Genes with more than 15% methylation change
Model 7	Significant genes with methylation level change more than 15% and obtained after combining methylation and expression significance values
Model 8	Significant genes with methylation level change more than 15%, inversely correlated with expression and obtained after combining methylation and expression significance values
Model 9	Significant genes with methylation level change more than 15%, positively correlated with expression and obtained after combining methylation and expression significance values
Model 10	Significant genes with methylation level change more than 40% and obtained after combining methylation and expression significance values

Table 3-1 Short summary of each analysis model. For the search of finding the optimal analysis strategy, we have applied 10 different analysis models on different data selection options.

Results regarding first four models are represented in Table 3-2 and next four models in Table 3-3. For each of these categories we have set four different result reporting options; only Batch230 results, only Batch250 results, combination of individual batch results (Pathways combined dataset) and Batch230+Batch250 (Pooled) dataset results.

	Model 1 (Differential Expression FDR<0.01)				**Model 2** (Differential Methylation FDR<0.01)				**Model 3** (Both Diff. and Meth Diff. Expressed FDR<0.01)				**Model 4** (Meth, Expr. Significances Combined FDR<0.01)			
KEGG TERMs	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
MAPK Signalling	**6**	**7**	**4**	**1**	**10**	**7**	**7**	**17**	**7**	-	**12**	**13**	**17**	**8**	**9**	**8**
ECM Receptor	**5**	**5**	**3**	**8**	26	-	52	**8**	**8**	-	**13**	**1**	**6**	**1**	**3**	**5**
ErbB Signalling	**13**	28	**20**	45	28	**1**	**6**	28	**14**	-	**20**	43	**20**	22	**20**	**7**
NF-KB Signalling	32	**11**	**17**	85	**11**	40	**18**	24	29	-	33	47	71	33	46	**17**
Wnt-β-Catenin Signalling	51	25	35	73	**14**	27	**16**	**13**	-	-	-	35	86	46	60	28
VEGF Signalling	46	47	47	105	42	**13**	23	85	36	-	42	-	85	84	82	56
Thyroid Cancer	30	52	34	66	55	42	43	-	**12**	**4**	**4**	65	88	65	74	69
Adherens Junction	34	**16**	**19**	24	21	**19**	**20**	**1**	34	**5**	**16**	**9**	21	**19**	**15**	**9**
p53 Signalling	**11**	**18**	**11**	**15**	68	-	77	59	**11**	-	**18**	23	34	**16**	22	30
TGF-beta Signalling	**3**	62	**12**	**5**	**18**	**6**	**12**	23	-	**1**	**5**	**5**	28	29	29	**19**
Notch Signalling	60	58	60	57	93	**18**	42	**6**	-	-	-	29	59	**9**	25	**13**
GnRH Signalling	61	27	42	26	32	24	24	40	31	-	38	26	73	41	54	68
Neurotrophin Signalling	**16**	**9**	**8**	**14**	**8**	**4**	**3**	30	**5**	-	**10**	24	**12**	**5**	**6**	**15**
Focal Adhesion	**8**	**1**	**2**	**7**	**1**	**2**	**1**	**2**	**3**	-	**3**	**2**	**7**	**2**	**2**	**3**
Transcr. Misregulation	42	**21**	29	65	**13**	-	40	-	-	-	-	37	31	23	28	34
Apoptosis	**17**	**19**	**16**	10	**12**	-	39	**20**	**15**	-	21	**16**	29	35	31	**6**
Pathways in Cancer	**1**	**2**	**1**	**3**	**2**	**3**	**2**	**16**	**1**	-	**1**	**3**	**1**	**2**	**1**	**2**
Toll-like receptor signalling pathway	**14**	57	24	32	67	-	76	60	28	-	32	-	102	85	92	43
Pentose-phosphate pathway	91	70	80	88	-	43	99	88	-	-	-	81	112	87	101	85

Table 3-2 Rankings of previously identified thyroid-cancer associated pathways in PANOGA functional enrichment results. For each analysis strategy four different results are shown in order to understand differences between different data selection strategies; A) Batch230 Results B) Batch250 Results C) Functional Enrichment Results of different Batches combined D) Batch230+Batch250 (Pooled) results. Model 1 represents, genes with differential expression FDR<0.01, Model 2 represents, genes with differential methylation FDR <0.01, Model 3 represents, genes having both differential expression and differential methylation FDR <0.01, Model 4 represents, genes that have FDR<0.01 after significance values of methylation and expression are combined. Bold numbers represent the genes in top 20 rankings.

	Model 5 (>15% Methylation Change and Inversely Correlated)				**Model 6** (>15% Methylation Change)				**Model 7** (>15% Methylation Change, Significances Combined)				**Model 8** (>15% Methylation Change, Inverse Correlated,Significances Combined)			
KEGG TERMs	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
MAPK Signalling	-	-	-	45	**16**	**3**	**2**	**2**	29	**4**	**8**	**6**	**-**	-	**-**	59
ECM Receptor	**14**	-	28	**6**	**8**	-	35	**3**	**3**	**10**	**3**	**3**	**18**	-	34	**7**
ErbB Signalling	-	**15**	**13**	**5**	26	**6**	**7**	**7**	28	**8**	**13**	**10**	31	**3**	**4**	**4**
NF-KB Signalling	22	-	35	29	**18**	-	41	31	**20**	**-**	39	22	**5**	-	**19**	**11**
Wnt-β-Catenin Signalling	21	-	34	72	64	28	40	46	51	**-**	62	72	-	-	**-**	-
VEGF Signalling	25	24	**17**	51	44	34	36	37	50	30	33	30	34	21	**17**	60
Thyroid Cancer	28	**14**	**15**	88	32	27	22	106	39	**20**	22	45	28	**17**	**13**	65
Adherens Junction	**9**	**8**	**5**	**7**	25	24	**15**	**6**	26	**7**	**9**	**14**	**8**	**-**	24	**10**
p53 Signalling	-	-	-	26	-	-	-	80	40	**-**	50	54	-	-	**-**	25
TGF-beta Signalling	-	-	-	**12**	41	32	32	**9**	**10**	**3**	**4**	**13**	**-**	-	**-**	-
Notch Signalling	**5**	**9**	**2**	49	-	**4**	26	24	54	**18**	28	**12**	**-**	22	50	55
GnRH Signalling	-	**6**	23	21	**19**	33	21	40	-	41	78	56	**2**	-	**8**	37
Neurotrophin Signalling	**7**	**2**	**3**	**15**	**15**	**14**	**8**	**12**	**13**	**13**	**7**	**17**	**20**	-	**36**	21
Focal Adhesion	**4**	-	**16**	**1**	**3**	**17**	**4**	**1**	**1**	**6**	**2**	**2**	**7**	-	23	**2**
Transcr. Misregulation	-	-	-	40	30	-	48	77	58	**-**	68	51	-	-	**-**	65
Apoptosis	**2**	-	**7**	**8**	**20**	-	42	**20**	**6**	**-**	27	**4**	25	-	38	**1**
Pathways in Cancer	**3**	-	**8**	**3**	**1**	**1**	**1**	**8**	**2**	**1**	**1**	**1**	**3**	-	**11**	**3**
Toll-like receptor signalling pathway	-	-	-	75	-	-	-	-	-	**-**	-	**8**	-	-	-	**5**
Pentose-phosphate pathway	**18**	**12**	**10**	25	-	30	64	-	34	31	29	**18**	**4**	-	**14**	**17**

Table 3-3 Rankings of previously identified thyroid-cancer associated pathways in PANOGA functional enrichment results - Part2. For each analysis strategy four different results are obtained in order to understand differences between different data selection strategies; A) Batch230 Results B) Batch250 Results C) Functional Enrichment Results of different Batches combined D) Batch230+Batch250 (Pooled) results. Model 5 represents, genes having more than 15% methylation change and are inversely correlated with expression values, Model 6 represents genes having more than 15% methylation change, Model 7 represents genes having more than 15% methylation change and having FDR<0.01 after significance values of methylation and expression are combined and finally Model 8 represents, genes having more than 15% methylation change, inversely correlated with expression values and having FDR<0.01 after significance values of methylation and expression are combined.

In the pooled dataset as threshold of 40% yielded highest ratio of inversely correlated genes, we have also made comparison between thresholds of 40% and 15% (Table 3-4). As a result, setting a methylation change threshold of 15% clearly outperformed setting a threshold of 40%.

	Model 7 (>15% Methylation Change, Significances Combined)	**Model 10** (>40% Methylation Change, Significances Combined)
KEGG TERMs	Pooled Dataset	Pooled Dataset
MAPK Signalling	**6**	-
ECM Receptor	**3**	**2**
ErbB Signalling	**10**	**10**
NF-KB Signalling	22	-
Wnt-β-Catenin Signalling	72	-
VEGF Signalling	30	24
Thyroid Cancer	45	-
Adherens Junction	**14**	**13**
p53 Signalling	54	-
TGF-beta Signalling	**13**	-
Notch Signalling	**12**	-
GnRH Signalling	56	**15**
Neurotrophin Signalling	**17**	**9**
Focal Adhesion	**2**	-
Transcr. Misregulation	51	-
Apoptosis	**4**	**1**
Pathways in Cancer	**1**	-
Toll-like receptor signalling pathway	**8**	-
Pentose-phosphate pathway	**18**	-

Table 3-4 Rankings of KEGG functional enrichment results on pooled dataset. After combining methylation and expression significances, genes having FDR<0.01 and having methylation change >15% and >40% are compared.

Moreover, we have compared the effects of inverse and positive correlation and whether which model informs more about the disease state (Table 3-5).

	Model 7 (>15% Methylation Change, Significances Combined)	**Model 8** (>15% Methylation Change, Inverse Correlated,Significances Combined)	**Model 9** (>15% Methylation Change, Positively Correlated,Significances Combined)
KEGG TERMs	Pooled Dataset	Pooled Dataset	Pooled Dataset
MAPK Signalling	**6**	59	**7**
ECM Receptor	**3**	**7**	**4**
ErbB Signalling	**10**	**4**	**5**
NF-KB Signalling	22	**11**	-
Wnt-β-Catenin Signalling	72	-	-
VEGF Signalling	30	60	53
Thyroid Cancer	45	65	27
Adherens Junction	**14**	**10**	**11**
p53 Signalling	54	25	-
TGF-beta Signalling	**13**	-	42
Notch Signalling	**12**	55	49
GnRH Signalling	56	37	50
Neurotrophin Signalling	**17**	21	**6**
Focal Adhesion	**2**	**2**	**1**
Transcr. Misregulation	51	65	-
Apoptosis	**4**	**1**	-
Pathways in Cancer	**1**	**3**	**3**
Toll-like receptor signalling pathway	**8**	**5**	-
Pentose-phosphate pathway	**18**	**17**	**15**

Table 3-5 Rankings of KEGG functional enrichment results on pooled dataset to investigate the differences between positive and inverse correlation. Bold numbers represent the genes in top 20 rankings.

When only inverse correlated genes were taken, we have observed 9 pathways in top 20 rankings (Model 8) and when only positively correlated genes were taken (Model 9), we have observed 8 pathways in top 20 rankings. On the other hand, when no filter applied and all genes above the 15% threshold were taken, we were able to reach the optimal analysis model with 12 pathways in top 20 (Model 7).

Overall, Model 7 was superior to other models at finding thyroid cancer related pathways in top 20 functional enrichment rankings. From this reason, identification of important transcription factors and more detailed functional enrichment analysis using ConsensusPathDB are conducted for the genes in Model 7.

3.1.5 Thyroid cancer - associated genes

We have investigated thyroid cancer-associated genes with respect to their methylation and expression significances in our datasets (Tables 3-6 - 3-8). Out of 25 thyroid-cancer associated genes retrieved from literature, for Batch230 there were a total of 6 differentially methylated and 13 differentially expressed genes whereas for Batch250 there were only two differentially methylated and nine differentially expressed genes with FDR<0.01. On the other hand, we observed a decent increase in the numbers of thyroid cancer-associated genes for the pooled dataset where 16 of the genes were found as differentially methylated and 19 as differentially expressed.

When significance values of differential methylation and differential expression were combined for each gene, we were able to capture two additional genes (SLC5A8 and NOTCH4) for Batch230 and one additional gene (RAP1GAP) for Batch250. Upon performing the same analysis for the pooled dataset, we observed 18 differentially altered genes, which was the highest compared to the previous dataset options. The results for the pooled dataset covered all of the genes that were captured on individual batch results, therefore besides combining significance values, pooling, i.e. expanding the dataset, aids at capturing disease-related genes with higher ratio.

	Batch230			
	DMR(FDR)	DE (FDR)	FDRs Combined	Methylation Change (percentage)
RAP1GAP	9.85E-04	4.42E-10	1.28E-11	**-15.75%**
TIMP3	-	-	-	**-28.98%**
DAPK	2.61E-03	2.38E-06	1.23E-07	**24.97%**
SLC5A8	-	-	1.03E-03	-3.96%
RARB	-	7.41E-03	2.67E-03	-3.22%
TSHR	2.98E-03	-	7.10E-03	-9.23%
RASSF6	2.59E-04	-	1.54E-03	**-37.42%**
CDKN2A	-	6.15E-05	1.22E-04	3.38%
MLH1	-	-	-	-1.34%
FN1	-	1.86E-09	1.27E-09	**-33.47%**
FOXE1	-	-	-	1.17%
HGF	-	-	-	-14.98%
KRT19	-	4.26E-11	1.64E-10	**-33.47%**
LGALS3	8.55E-03	1.27E-13	3.85E-14	**-11.78%**
MET	-	1.13E-17	7.35E-18	**-46.44%**
RET	-	1.22E-03	4.33E-04	**-18.49%**
KISS1R	-	2.84E-05	3.77E-05	-4.81%
ADAMTS5	-	3.53E-03	8.71E-04	**27.62%**
HOXB4	-	-	-	-3.01%
TCL1B	3.50E-03	-	-	**-35.18%**
NOTCH4	-	-	9.04E-03	-12.10%
RASSF1	-	-	-	**16.56%**
PPARG	-	3.04E-03	1.21E-03	-2.92%
ALK	-	3.27E-09	9.25E-10	-2.57%
NTRK3	-	-	-	-2.15%

Table 3-6 Batch230 Results showing Differential Methylation (DMR), Differential Expression (DE), Combination of Differential Methylation and Differential Expression Significances (FDRs Combined) and Methylation Change in %. Only the values with Differential Expression and Differential Methylation Significances below 0.01 are shown on the table. FDR>0.01 are shown as blank. Moreover, methylation changes >15% are shown as bold.

	Batch250				
	DMR(FDR)	DE (FDR)	FDRs Combined	Methylation (percentage)	Change
RAP1GAP	-	-	8.31E-03	**-18.04%**	
TIMP3	-	-	-	**-36.24%**	
DAPK	-	4.38E-06	2.60E-06	**-23.55%**	
SLC5A8	-	-	-	-11.37%	
RARB	-	9.34E-04	5.75E-04	**-23.60%**	
TSHR	-	-	-	**-17.94%**	
RASSF6	-	-	-	-14.54%	
CDKN2A	-	1.58E-07	1.65E-06	14.95%	
MLH1	-	-	-	**-18.74%**	
FN1	-	6.59E-10	3.45E-10	**-39.00%**	
FOXE1	-	-	-	2.18%	
HGF	-	-	-	-11.01%	
KRT19	-	9.45E-09	1.17E-08	**-27.31%**	
LGALS3	-	2.58E-07	1.21E-07	-2.92%	
MET	-	1.10E-09	2.25E-09	**-44.72%**	
RET	-	-	-	-12.35%	
KISS1R	-	1.64E-04	4.23E-04	-1.33%	
ADAMTS5	-	-	-	**30.39%**	
HOXB4	-	-	-	-4.06%	
TCL1B	8.97E-04	-	7.54E-04	**-37.49%**	
NOTCH4	-	-	-	**-22.75%**	
RASSF1	5.14E-03	-	-	**20.21%**	
PPARG	-	-	-	7.98%	
ALK	-	1.05E-03	8.55E-04	**-19.68%**	
NTRK3	-	-	-	-4.02%	

Table 3-7 Batch250 Results showing Differential Methylation (DMR), Differential Expression (DE), Combination of Differential Methylation and Differential Expression Significances (FDRs Combined) and Methylation Change in %. Only the values with Differential Expression and Differential Methylation Significances below 0.01 are shown on the table. FDR>0.01 are shown as blank. Moreover, methylation changes >15% are shown as bold.

	Pooled Dataset			
	DMR(FDR)	DE (FDR)	FDRs Combined	Methylation Change (percentage)
RAP1GAP	1.29E-05	4.42E-10	1.93E-13	**-32.37%**
TIMP3	-	1.46E-03	-	**-31.69%**
DAPK	2.14E-04	1.26E-11	3.35E-12	**24.69%**
SLC5A8	-	4.80E-04	1.99E-04	-11.61%
RARB	7.43E-03	6.77E-07	1.01E-07	**-17.92%**
TSHR	6.58E-06	1.05E-07	8.96E-06	**-18.12%**
RASSF6	-	1.05E-07	2.31E-08	3.85%
CDKN2A	-	1.57E-11	4.87E-11	2.92%
MLH1	-	-	-	-1.10%
FN1	7.00E-04	3.78E-16	1.16E-17	**-39.43%**
FOXE1	-	-	-	1.00%
HGF	-	7.65E-03	-	-1.54%
KRT19	2.36E-03	2.83E-18	3.17E-19	-9.87%
LGALS3	9.21E-05	1.46E-19	7.22E-22	-9.98%
MET	1.09E-04	1.99E-26	1.50E-28	**-45.69%**
RET	3.08E-03	1.01E-04	4.96E-06	**-15.86%**
KISS1R	1.08E-03	5.98E-11	2.02E-12	-4.12%
ADAMTS5	7.80E-05	-	8.94E-03	**28.55%**
HOXB4	1.07E-03	-	7.07E-03	-3.21%
TCL1B	5.29E-03	-	1.55E-06	**-35.67%**
NOTCH4	6.85E-03	4.37E-04	4.10E-05	-13.05%
RASSF1	1.62E-03	-	5.08E-03	**19.06%**
PPARG	-	1.41E-05	7.76E-06	-2.22%
ALK	4.69E-04	3.79E-13	6.63E-15	9.44%
NTRK3	-	4.74E-03	4.15E-03	-1.44%

Table 3-8 Pooled Dataset Results showing Differential Methylation (DMR), Differential Expression (DE), Combination of Differential Methylation and Differential Expression Significances (FDRs Combined) and Methylation Change in %. Only the values with Differential Expression and Differential Methylation Significances below 0.01 are shown on the table. FDR>0.01 are shown as blank. Moreover, methylation changes >15% are shown as bold.

3.2 Discussion

For the purpose of understanding the interplay between expression and methylation in thyroid cancer, we have conducted comparisons between four data and ten analysis strategies with respect to the observance rate of thyroid related pathways in the functional enrichment results (Tables 3-2 – 3-5). Moreover, we have also conducted a threshold analysis to understand whether setting a methylation change threshold improves the outcome of the experiment.

3.2.1 Methylation threshold analysis

In order to identify the benefits of setting a methylation level threshold, we have conducted a threshold analysis for various threshold levels by calculating the inverse correlation ratio between methylation and expression. When only inverse correlation ratios above different thresholds were looked at, best performing threshold was 35% for both Batch230 and Batch250 and 40% for the pooled dataset (Figure 3-1). However, the reason behind setting a threshold is to witness a concrete difference between above and below thresholds. In this sense, optimal threshold would be the one that maximizes the difference between ratio above and below of a certain threshold.

When investigating the total inverse correlation gain for all three datasets, best performing threshold level was found at "15%" with 29.77% correlation gain where improvement in inverse correlation between change in methylation level and expression reached its highest value (Figure 3-2).

Consequently, when 15% methylation change threshold was added to Model 4, which previously possessed maximum number of thyroid-cancer associated pathways in top20 functional enrichment rankings, we were able to reach the optimal analysis strategy with 12 thyroid-cancer associated pathways in top 20 rankings (Model 7) (Tables 3-2 – 3-5). Similarly, when Model 2 and Model 6 were compared to each other, addition of 15% methylation change threshold improved the functional enrichment results by additionally identifying ErbB signalling, TGF-beta signalling and Neurotrophin signalling pathways in top 20 rankings. Thus, it can be argued that the genes with more than 15% methylation change may be the core reason behind changes in these pathways, which were all associated with thyroid-cancer in previous works.

Moreover, we have also compared functional enrichment results between Model 7, 15% methylation threshold and Model 10, 40% methylation threshold, which did not have the highest correlation gain but had the highest inverse correlation percentage in the pooled dataset. As a result, setting 15% threshold level clearly outperformed threshold of 40% (Table 3-4), implying that the information of a "gain of inverse correlation" above and below the threshold is more important than "overall inverse correlation" ratio above the threshold.

3.2.2 Combining methylation and expression data

Due to the reason that methylation and gene expression have different roles in the development of thyroid cancer, combining significance values obtained from methylation and expression studies leads to a better detection of thyroid-related genes (Tables 3-6 – 3-8). To exemplify, for Batch230, SLC5A8 gene was not detected as significantly expressed or significantly methylated. However, when the significances of expression and methylation were combined, we observed SLC5A8 as significantly altered with false discovery rate of 0.001. Similar cases were also observed for Batch250 and pooled dataset, hence combining methylation and gene expression information on pooled data enabled us to obtain highest ratio (21 out of 25) of detecting thyroid-cancer associated genes as significantly altered.

Moreover, for the purpose of understanding the reflection of combining methylation and expression significances on functional enrichment results, we have compared Model 6 (>15% methylation change) with Model 7 (>15% methylation change and methylation, expression significances combined) and Model 4 (Only methylation, expression significances combined) with Model 1 (Only differential expression) and Model 2 (Only differential methylation).

Considering the pooled dataset, for Model 6, we have observed 9 thyroid-cancer associated pathways in top20 functional enrichment results whereas for Model 7, which is the same dataset with only methylation and expression significances were combined, we have detected 12 thyroid-cancer associated pathways in top20 functional enrichment results. Similarly, for Model 1 there were 7 and for Model 2 there were 8 thyroid-cancer associated pathways in top20 functional enrichment results. When expression and methylation significances were combined instead of treated separately, we were able to observe 11 important pathways in top 20 functional enrichment results (Model 4). Moreover, there were various pathways that were not captured at all in Model 1 and 2, which were only captured when the significances of expression and methylation were combined. For example; in Batch250 differential methylation functional enrichment results (Model 2, B dataset), p53-

signalling pathway was not listed as significant at all, with Bonferroni Score above 0.01. When methylation and expression significances were combined (Model 4, B dataset), p53 signalling pathway was observed at 16th rank with Bonferroni Score 1.51E-12. Similar improvement was also observed between Model 5 and Model 8, as combining significances led to an improved performance with additional detection of toll-like receptor pathway in top 10 rankings.

Consequently, incorporating methylation and expression information together not only improved detection rate of disease-specific genes but it also increased the rankings of disease-specific pathways in functional enrichment results.

Overall, when the data was pooled, methylation, expression significances were combined and only genes with more than 15% methylation change were selected, best performing results were reached with 12 pathways in top20 functional enrichment results (Table 3-2 – 3-5) namely; MAPK signalling, Extracellular matrix receptor, ErbB signalling, TGF-beta signalling, Notch signalling, Neurotrophin signalling, Apoptosis, Focal adhesion, Pathways in cancer, Toll-like receptor signalling, Pentose-phosphate and Adherens junction pathways.

3.2.3 Testing on an independent dataset

In addition to the supporting articles from the literature, for the purpose of proving the generalizability and efficiency of our proposed framework, we have applied the same procedures described above on another independent dataset with 30 samples retrieved from thyroid cancer experiments in TCGA. To achieve that, firstly we have calculated the methylation threshold value with "maximum inverse correlation gain", which was also 15% for the test dataset and secondly, we have combined methylation and expression significances by using Fisher's weighted Z-method. As a result, compared to our training dataset results, we were able to obtain similar pathways in similar rankings in the test dataset, hence there were 11 thyroid cancer-associated pathways in top 20 functional enrichment rankings (Table 3-9). These findings also support that our approach can be applied to different, independent cancer datasets, which may aid at detecting important pathways for other cancer types as well.

	Training Dataset (>15% Methylation Change, Significances Combined)	**Test Dataset** (<15% Methylation Change, Significances Combined)
KEGG TERMs	Pooled	Pooled
MAPK Signalling	**6**	**11**
ECM Receptor	**3**	**1**
ErbB Signalling	**10**	**16**
NF-KB Signalling	22	25
Wnt-β-Catenin Signalling	72	80
VEGF Signalling	30	69
Thyroid Cancer	45	68
Adherens Junction	**14**	**19**
p53 Signalling	54	**17**
TGF-beta Signalling	**13**	**7**
Notch Signalling	**12**	83
GnRH Signalling	56	**15**
Neurotrophin Signalling	**17**	**12**
Focal Adhesion	**2**	**2**
Transcr. Misregulation	51	-
Apoptosis	**4**	**20**
Pathways in Cancer	**1**	**4**
Toll-like receptor signalling pathway	**8**	-
Pentose-phosphate pathway	**18**	75

Table 3-9 Comparison between the training and test dataset.

3.2.4 Disease etiology

Although there may be other mechanisms at play leading to the thyroid cancer state, in this work we have mainly investigated pathways, which were mainly influenced by expression changes highly correlated with methylation changes. While searching for the optimal model, several common pathways were observed at different rankings in almost all of the models, reassuring that methylation change might disturb certain pathways that might be involved in thyroid cancer etiology. When focusing only on differential methylation results, we have observed significant changes in important pathways such as MAPK Signalling, Wnt-β-catenin Signalling, Notch signalling, Apoptosis and TGF-beta signalling pathways. Besides the pathways that were directly affected by methylation, other secondary molecular mechanisms were also triggered, such as Transcriptional misregulation, Thyroid cancer and p53 signalling pathways, which were only captured by expression experiments. Specifically, when pooled data results, which possess more than 40% methylation change, were being

investigated, we observed significant changes in Apoptosis, Extracellular matrix, ErbB, VEGF, GnRH and Neurotrophin signalling pathways. Thus, it is more probable that the core reason behind major changes in these pathways may be due to high methylation level change between disease and normal state (Table 3-4).

In our analysis, optimal analysis strategy which yielded maximum number of thyroid-cancer associated pathways in top rankings was found to be Model 7. When the functional enrichment results of the best-performing analysis model was investigated in detail, all of the top20 ranked pathways on the list could be associated with thyroid cancer**.** In addition to the thyroid-cancer related pathways that were extracted from literature at the beginning, Endocytosis (Lanzetti & Di Fiore, 2008), Glutamate (Rzeski, Ikonomidou, & Turski, 2002), Proteasome (Conticello et al., 2007), Gluconeogenesis and glycolysis (Shulman, Ladenson, Wolfe, Ridgway, & Wolfe, 1985) pathways are found as linked to thyroid cancer in previous works about thyroid cancer.

Furthermore, when the details of 2826 genes that have >15% methylation change were explored, some of the GO: Biological Process terms with high significance were: regulation of signal transduction, cell differentiation, phosphate containing metabolic process, morphogenesis and neuron development. For each annotation term, we have performed KEGG functional analysis to examine the association with the cancer state (Table 3-10). Almost all of the terms were found to be associated with “Pathways in Cancer” which was also supported by the recent literature works (Dvorak, 2002; Hanahan & Weinberg, 2000; Klein & Ojamaa, 2001; Kouvaraki et al., 2005; Owens et al., 1996; Ozcan et al., 2011; Saharinen, Tammela, Karkkainen, & Alitalo, 2004; Sherr, 2000; Vivanco & Sawyers, 2002; J. Yang & Weinberg, 2008).

GO: Biological Process Terms	No. of genes that overlap with our list	Associated q-value	No. of genes in cancer pathway	Association with cancer pathway
Regulation of Signal Transduction	540 (21.5%)	1.54E-29	56	2.06E-15
Cellular Development Process, Cell Differentiation	633 (19.2%)	2.16E-21	74	4.02E-23
Phosphate Containing Compound Metabolic Process	653 (19.1%)	2.16E-21	59	2.35E-14
Anatomical Structure Formation, Morphogenesis	271 (22.7%)	8.04E-17	34	1.72E-11
Neuron Development	229 (23.9%)	8.47E-17	29	1.57E-09
Actin Cytoskeleton Organization	142 (27.6%)	1.53E-15	13	1.46E-03
Regulation of Catalytic Activity	411 (19.8%)	3.30E-15	48	5.03E-15
Circulatory System Development	199 (23.3%)	1.67E-13	40	1.45E-19
Cell Junction Assembly	70 (33.8%)	3.87E-12	10	3.12E-04
Vasculature Development	139 (24.9%)	1.28E-11	28	1.27E-13
Regulation of Adhesion	137 (24.7%)	2.71E-11	25	1.46E-11
Regulation of Programmed Cell Death	339 (19.3%)	5.69E-11	48	2.25E-17
Protein Kinase Activity	180 (22.0%)	4.18E-10	30	3.16E-15
Response to External Stimulus	142 (23.1%)	1.83E-09	20	4.95E-07
Epithelium Development	214 (20.5%)	4.47E-09	46	1.07E-25
Response to Growth Factor	151 (22.1%)	9.85E-09	35	5.21E-19
Protein Modification Process	526 (17.0%)	5.44E-08	57	6.96E-17
Regulation of Developmental Process	190 (19.6%)	1.19E-06	36	7.49E-17
Regulation of Cell Growth	71 (21.5%)	4.62E-05	9	3.88E-03
Mesonephros Development	27 (26.0%)	3.82E-04	12	2.08E-10

Table 3-10 GO: Biological Process annotation table for significantly altered genes in Model 7 obtained using ConsensusPathDB. Out of 340 GO: Biological Process terms with q-value <0.01, information of 20 important terms are reported. For each annotation term in the list, we have conducted KEGG Pathway Analysis. Almost all of the terms were significantly associated with "Pathways in Cancer".

Moreover, since post-translational modification and regulation of transcription pathways are critical for cancer diagnosis and therapy (Darnell, 2002; Krueger & Srivastava, 2006), we have searched for transcription factors in TFCat database (Fulton, et al., 2009) and as a result, 207 out of 2826 genes (7.32%) were annotated as transcription factors and 245 out of 2826 genes (8.66%) were annotated as being involved in post-translational modification processes with Benjamini significance of "7.98E-05" in ConsensusPathDB analysis. Consequently, these genes may be active at altering other pathways, revealing other mechanisms involved in thyroid cancer.

4. INVESTIGATING THE INTERPLAY BETWEEN METHYLATION AND EXPRESSION IN DIFFERENT CANCER TYPES

4.1 Results

In order to extract insights of cancer disease aetiologies, firstly we have combined differential expression and differential methylation significances for each gene as an alternative approach to individual expression or individual methylation analyses. Moreover, we have only considered the genes having combined methylation, expression significance <0.01 which were merged using Fisher's Z method. As a result, there were high numbers of differentially altered genes for each cancer type, which was supporting our intuition to set a methylation level change threshold. Specifically, in thyroid cancer data we have found 4102 differentially altered genes and in breast cancer data we have found 7529 significantly altered genes. Additionally, we have found 6681 differentially altered genes in colon cancer data whereas in prostate cancer data we have found 5674 significantly altered genes.

4.1.1 Identifying important genomic regions

Using ChAMP, we were able to obtain methylation change information for six different regions of a gene. One of our main focuses was whether certain genomic regions look more promising compared to other genomic regions. In this sense, we have compared average inverse correlation ratios for the regions with similar features (Table 4-1). In more detail, average inverse correlation ratio of 1stExon regions (73.89%) was higher than Body regions (57.82%). Similarly, 5'UTR regions (68.45%) had higher inverse correlation ratio compared to 3'UTR regions (52.25%) and TSS200 regions (76.47%) had higher inverse correlation ratio compared to TSS1500 regions (67.97%). Therefore, with the idea of higher inverse correlation would inform more about the disease mechanisms, we have selected 1stExon, TSS200 and 5UTR at further analysis.

Threshold Levels	**1stExon**	3UTR	**5UTR**	Body	TSS1500	**TSS200**
50%	47.55%	3.13%	44.76%	38.13%	71.31%	36.33%
45%	61.14%	30.56%	54.80%	38.47%	58.54%	61.96%
40%	57.09%	30.99%	56.48%	64.68%	78.80%	87.85%
35%	76.80%	58.11%	71.95%	59.65%	75.62%	80.70%
30%	75.03%	54.71%	72.18%	59.62%	73.10%	82.20%
25%	78.14%	52.41%	71.30%	58.90%	70.31%	80.20%
20%	74.64%	50.64%	69.14%	57.85%	67.30%	77.39%
15%	73.00%	49.84%	66.34%	57.08%	64.50%	73.71%
10%	70.44%	50.38%	64.22%	56.30%	62.80%	71.13%
5%	69.19%	49.67%	64.04%	55.36%	62.15%	69.96%
	73.89%	52.25%	**68.45%**	57.82%	67.97%	**76.47%**

Table 4-1 Comparison of average inverse correlation ratios between different genomic regions to decide on which genomic regions are of higher importance. Only genes having differential methylation Bonferroni score <0.05 are selected for further analysis. Moreover, threshold analysis is also conducted for varying thresholds between 5-50% and average correlation ratios are stated at the bottom of the table.

4.1.2 Methylation threshold analysis

After identifying the important genomic regions, we have focused on setting a valid methylation threshold for our analysis by checking the inverse correlation between differentially methylated ($p<0.05$) and differentially expressed ($p<0.05$) genes. Moreover, we have performed threshold analysis for varying thresholds between 5% and 50%.

As a result of our analysis, threshold of 25% have led us to highest average inverse correlation when only 5'UTR, 1st Exon and TSS200 regions were selected. (Table 4-2). From that point on, we have used the reduced dataset, which had genes having significant expression, methylation change (combined p-value <0.01) and which have methylation change larger than 25%. In total, there were 196 differentially altered genes in thyroid cancer dataset whereas in breast cancer dataset, there were 683 differentially altered genes.

Additionally, there were 1002 differentially altered genes in colon cancer data and there were 348 differentially altered genes in total for prostate cancer data.

Meth. Change	1stExon	5UTR	TSS200	Average
50%	47,55%	44,76%	36,33%	42,88%
45%	61,14%	54,80%	61,96%	59,30%
40%	57,09%	56,48%	**87,85%**	67,14%
35%	76,80%	71,95%	80,70%	76,48%
30%	75,03%	**72,18%**	82,20%	76,47%
25%	**78,14%**	71,30%	80,20%	**76,55%**
20%	74,64%	69,14%	77,39%	73,72%
15%	73,00%	66,34%	73,71%	71,02%
10%	70,44%	64,22%	71,13%	68,59%
5%	69,19%	64,04%	69,96%	67,73%

Table 4-2 Comparison of inverse correlation ratios for varying threshold levels.

4.1.3 Exploring commonly altered genes

One of our major goals was to detect commonly altered genes and pathways observed at different cancer types. A Venn diagram showing number of differentially altered genes are represented at Figure 4-1. 1 gene that was showing significance and shared among all cancer types was GPR115, which is a G protein-coupled receptor. Calculating probability of obtaining an overlap by randomly picking 196, 683, 1002 and 348 genes for BRCA, THCA, COAD and PRAD respectively also supported significance of this finding, as the procedure is repeated 10000 times and as a result, 54 times there was an overlap between four cancer types, which was corresponding to ratio of 0.005.

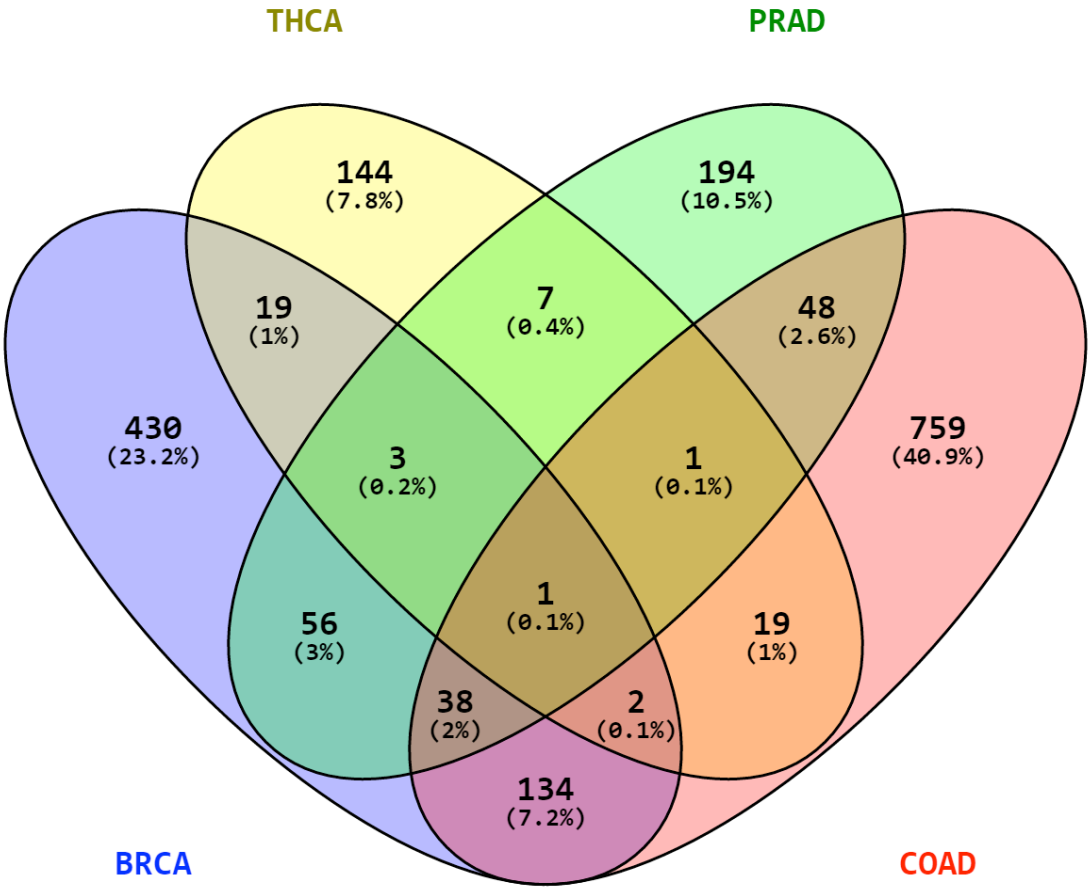

Figure 4-1 Venn diagram of significant genes, informing about on which datasets those genes are detected as significantly altered. Only genes having combined expression, methylation significance <0.01 are included at this analysis.

4.1.4 Functional enrichment – Extracting commonly altered pathways

In order to better understand disease aetiologies, functional enrichment analysis results of four different cancer types are revealed at Tables 4-3 - 4-6. As a result, there were 44, 21, 27 and 19 significantly altered pathways for BRCA, THCA, COAD and PRAD datasets, respectively. Thyroid cancer results involving top 15 Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2015) pathways are shown at Table 4-3. Breast cancer results are presented at Table 4-4 whereas colon cancer results are shown at Table 4-5. Finally, prostate cancer results are shown at Table 4-6.

KEGG ID	KEGG Term	Bonf-P-Value
KEGG:04610	Complement and coagulation cascades	1,90E-13
KEGG:04510	Focal adhesion	5,30E-12
KEGG:04012	ErbB signaling pathway	1,03E-09
KEGG:05220	Chronic myeloid leukemia	1,41E-09
KEGG:05221	Acute myeloid leukemia	1,61E-08
KEGG:04722	Neurotrophin signaling pathway	1,84E-08
KEGG:05210	Colorectal cancer	4,74E-08
KEGG:05100	Bacterial invasion of epithelial cells	2,31E-06
KEGG:04520	Adherens junction	3,48E-06
KEGG:04660	T cell receptor signaling pathway	1,77E-05
KEGG:05215	Prostate cancer	2,31E-05
KEGG:05213	Endometrial cancer	2,59E-05
KEGG:05143	African trypanosomiasis	3,42E-05
KEGG:05223	Non-small cell lung cancer	3,49E-05
KEGG:05216	Thyroid cancer	9,90E-05

Table 4-3 Top 15 Panoga functional enrichment results for the Thyroid Cancer Dataset (THCA). All significance values are stated as corrected Bonferroni P scores, whereas KEGG ids and explanation of the terms are included at the table as well.

KEGG ID	KEGG Term	Bonf-P-Value
KEGG:04721	Synaptic vesicle cycle	1,04E-12
KEGG:04610	Complement and coagulation cascades	2,70E-12
KEGG:04012	ErbB signaling pathway	7,75E-12
KEGG:04130	SNARE interactions in vesicular transport	1,89E-10
KEGG:04512	ECM-receptor interaction	7,38E-10
KEGG:03050	Proteasome	1,04E-09
KEGG:05140	Leishmaniasis	3,57E-09
KEGG:05133	Pertussis	3,57E-09
KEGG:00051	Fructose and mannose metabolism	3,77E-09
KEGG:05144	Malaria	8,60E-09
KEGG:04510	Focal adhesion	9,32E-09
KEGG:04810	Regulation of actin cytoskeleton	2,85E-08
KEGG:05214	Glioma	5,49E-08
KEGG:04650	Natural killer cell mediated cytotoxicity	2,10E-07
KEGG:05134	Legionellosis	5,95E-07

Table 4-4 Top 15 Panoga functional enrichment results for the Breast Cancer Dataset (BRCA). All significance values are stated as corrected Bonferroni P scores, whereas KEGG ids and explanation of the terms are included at the table as well.

KEGG ID	KEGG Term	Bonf-P-Value
KEGG:04020	Calcium signaling pathway	4,22E-15
KEGG:04512	ECM-receptor interaction	1,20E-13
KEGG:04610	Complement and coagulation cascades	7,84E-13
KEGG:04724	Glutamatergic synapse	6,76E-12
KEGG:04723	Retrograde endocannabinoid signaling	6,98E-12
KEGG:04514	Cell adhesion molecules (CAMs)	7,49E-12
KEGG:04080	Neuroactive ligand-receptor interaction	5,75E-10
KEGG:04330	Notch signaling pathway	1,53E-08
KEGG:04810	Regulation of actin cytoskeleton	3,40E-08
KEGG:04720	Long-term potentiation	7,18E-08
KEGG:04971	Gastric acid secretion	1,26E-07
KEGG:04510	Focal adhesion	2,64E-07
KEGG:04012	ErbB signaling pathway	4,42E-07
KEGG:05031	Amphetamine addiction	6,54E-07
KEGG:04722	Neurotrophin signaling pathway	1,72E-06

Table 4-5 Top 15 Panoga functional enrichment results for the Colon Cancer Dataset (COAD). All significance values are stated as corrected Bonferroni P scores, whereas KEGG ids and explanation of the terms are included at the table as well.

KEGG ID	KEGG Term	Bonf-P-Value
KEGG:04512	ECM-receptor interaction	4,69E-11
KEGG:04012	ErbB signaling pathway	9,74E-09
KEGG:00030	Pentose phosphate pathway	6,89E-08
KEGG:04660	T cell receptor signaling pathway	1,48E-07
KEGG:00051	Fructose and mannose metabolism	1,97E-07
KEGG:04710	Circadian rhythm	7,69E-07
KEGG:05215	Prostate cancer	3,04E-06
KEGG:00052	Galactose metabolism	1,52E-05
KEGG:00061	Fatty acid biosynthesis	1,63E-05
KEGG:05223	Non-small cell lung cancer	2,09E-05
KEGG:04666	Fc gamma R-mediated phagocytosis	2,95E-05
KEGG:00430	Taurine and hypotaurine metabolism	3,50E-05
KEGG:04664	Fc epsilon RI signaling pathway	5,54E-05
KEGG:05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	5,82E-05
KEGG:05220	Chronic myeloid leukemia	2,10E-04

Table 4-6 Top 15 Panoga functional enrichment results for the Prostate Cancer Dataset (PRAD). All significance values are stated as corrected Bonferroni P scores, whereas KEGG ids and explanation of the terms are included at the table as well.

Moreover, Venn diagram showing significantly altered pathways and the relation between the cancers are revealed at Figure 4-2. In addition, numbers of KEGG pathways that are shared among at least three cancer types are also presented at Table 4-7. Remarkably, detailed investigation showed us that, erbB signalling pathway is detected as being the most important pathway at our analysis since it was found as significantly affected for all cancer types at our analysis.

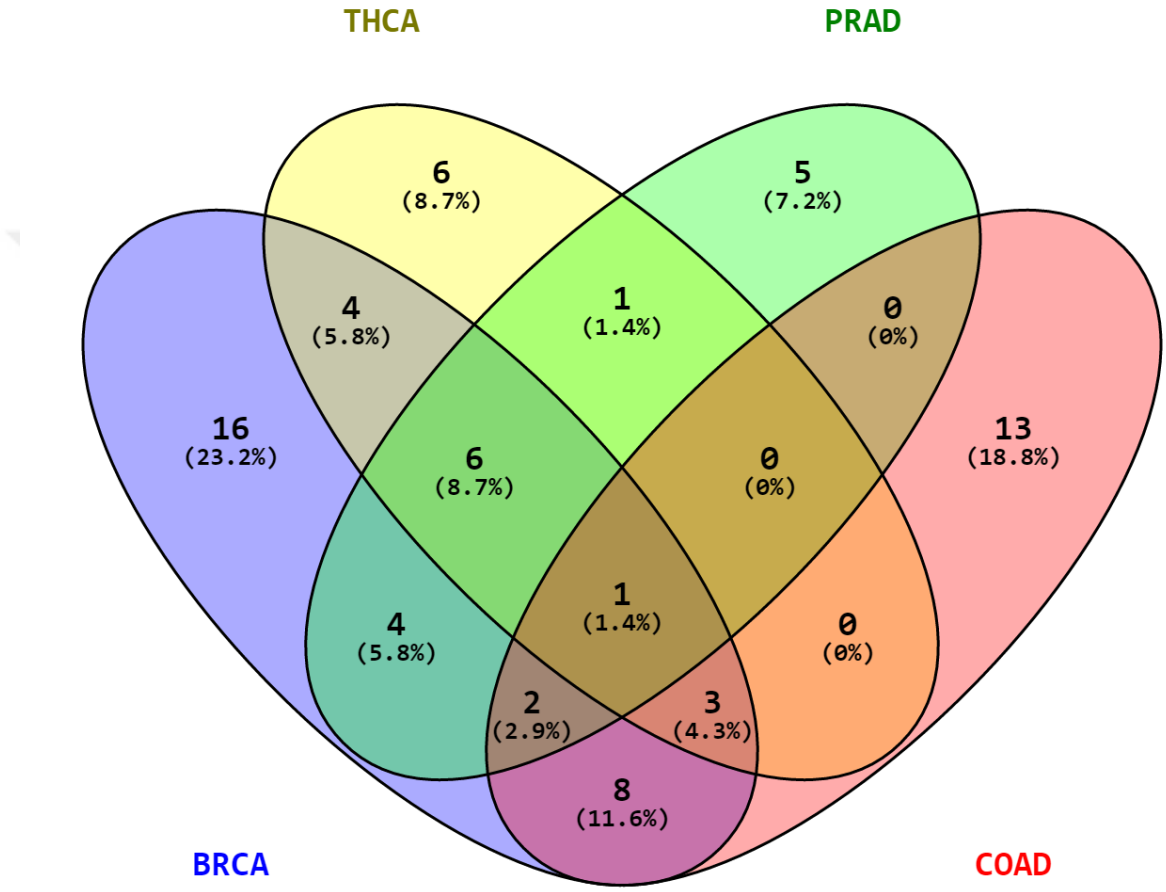

Figure 4-2 Venn diagram of significantly altered pathways informing about on which datasets those pathways are detected as significantly altered (Bonf.Score <0.01).

Pathway Name	Breast	Thyroid	Prostate	Colon
ErbB signaling pathway	**7.75E-12**	**1.03E-09**	**9.74E-09**	**4.42E-07**
Complement and coagulation cascades	**2.70E-12**	**1.90E-13**	**-**	**7.84E-13**
ECM-receptor interaction	**7.38E-10**	**-**	**4.69E-11**	**1.20E-13**
Focal adhesion	**9.32E-09**	**5.30E-12**	**-**	**2.64E-07**
Chronic myeloid leukemia	**8.94E-06**	**1.41E-09**	**2.10E-04**	**-**
Neurotrophin signaling pathway	**1.37E-05**	**1.84E-08**	**-**	**1.72E-06**
T cell receptor signaling pathway	**1.11E-06**	**1.77E-05**	**1.48E-07**	**-**
Glioma	**5.49E-08**	**1.47E-04**	**7.90E-04**	**-**
Prostate cancer	**2.69E-05**	**2.31E-05**	**3.04E-06**	**-**
Bacterial invasion of epithelial cells	**1.22E-05**	**2.31E-06**	**3.51E-04**	**-**
Non-small cell lung cancer	**4.71E-04**	**3.49E-05**	**6.57E-04**	**-**
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	**2.63E-04**	**-**	**5.82E-05**	**1.73E-04**

Table 4-7 List of significantly altered (Bonferroni score <0.01) pathways that are shared by at least three types of cancers and their Bonferroni scores associated with each dataset.

4.2 Discussion

In our study we show that methylation in 1stExon, 5'UTR and TSS200 are of higher importance since they show high inverse correlation with expression information, hence they may be more informative regarding disease status. Although there were a few numbers of studies showing direct correlation between methylation and expression especially at gene body regions, in our comparative study, we have evaluated the relations at different genomic regions considering the "body" regions as well. With the idea of inverse correlation has higher impact on gene expression, at further analysis, we have favoured the regions showing higher inverse correlation and we have only included the genes having significant methylation change at their 1st exon, TSS200 and 5'UTR regions. Moreover, by using 25% methylation change threshold, we were able to identify key genes and pathways that were mainly altered by changes in methylation levels contributing to disease state.

Our findings in our study regarding more important genomic regions are supported by various literature works, which also include wet lab procedures. To exemplify; Brenet et al. have shown at their work that "DNA Methylation at first exon is tightly linked to transcriptional silencing"(Brenet et al., 2011). Additionally, Laurent et al. have presented in their work that in human embryonic stem cells, methylation of CpGs within 1kb of transcription start site was found to be negatively correlated to gene expression, whereas methylation of CpGs in the gene body was found as positively correlated to expression (Laurent et al., 2010). In another work of Fleischer et al., 10 genes out of 27 gene promoters are shown to have significant correlation between methylation and expression, while all associations showed a negative correlation (Fleischer et al., 2014).

In order to validate our findings that we have found using our data integration method, we have searched for support from the literature. For thyroid cancer data (Table 4-3), acute myeloid leukaemia was found at most significantly altered pathway in our analysis. It has been previously shown that radiodine treatment for thyroid cancer triggers acute myeloid leukemia (Gilabert & Prebet, 2012) and the patients in our dataset may have experienced radiodine treatment for their thyroid cancer. Pathways such as adherens junction, complement coagulation cascade, erbB signalling, focal adhesion, neurotrophin signalling and VEGF signalling pathways were previously associated with thyroid cancer (Lucas et al., 1996; McGregor et al., 1999; Mohammadi-asl et al., 2011; Owens, et al., 1996; Vieira et al., 2005). Remarkably, other cancer pathways such as colorectal, bladder, prostate, endometrial and non-small cell cancer pathways were also detected at functional enrichment, which implies genes leading to thyroid cancer may also be involved in many cancer types or the cancer in thyroid tissue may have spread to other tissues, leading to significant functional enrichment results for other cancer types as well.

When looked at the functional enrichment results of breast cancer data (Table 4-4), 13 out of 15 pathways were linked to breast cancer previously in the literature (Almond & Cohen, 2002; Ames, Hallett, & Murphy, 2009; De Spiegeleer et al., 2015; J. Li et al., 1997; Lu, Weaver, & Werb, 2012; Niculescu, Rus, Retegan, & Vlaicu, 1992; Stern, 2000; Xu et al., 2000). Specifically, the terms leishmaniasis,pertussis, malaria and legionellosis were found as significantly altered at our analysis, which implies methylation-affected genes that are found in breast cancer, acting in a similar way to the genes that are found in bacterial invasion (De Spiegeleer, et al., 2015). This was also supported by various recent studies to test malaria drugs on metastatic breast cancers.

For the colon cancer data (Table 4-5), 14 out of 15 pathways were found related to colon cancer biology when looked at the literature (Agrez, 1996; Chang et al., 2005; Denlinger & Barsevick, 2009; Herath & Boyd, 2010; Joshi & Gardner, 1996; Kou, Zhang, Chen, & Hu, 2015; Najdi, Holcombe, & Waterman, 2011; Patsos et al., 2010; Prickett & Samuels, 2012; Sanz-Pamplona et al., 2014; Szmida et al., 2015; Tian, Sun, Zhao, Xiong, & Fang, 2015; Y. Wu et al., 2012) with the exception of amphetamine addiction which requires further studies to associate with colon caner. On the other hand for the prostate cancer data (Table 4-6), out of top 15 pathways that were found by our method, 14 were found as being linked to prostate cancer when looked at the previous studies (Britten, 2004; Jia, Liu, & Zhao, 2012; Jung-Hynes, Huang, Reiter, & Ahmad, 2010; J. Li, et al., 1997; Priolo et al., 2014; Stewart, Cooper, & Sikes, 2004; Swinnen et al., 2000; Tang et al., 2015; Tsouko et al., 2014). Specifically, T-cell coinhibition in prostate cancer is previously reported as evasion mechanism that is also targeted as therapeutic strategy. It is highly likely that the patients in our dataset may have T-cell inhibition to evade the cancer status as well (Barach, Lee, & Zang, 2011).

Additionally, we have also investigated individual methylation and individual expression results. In this sense, when only the genes having methylation level change above 25% and shared by at least 3 cancer types are selected for functional enrichment analysis, we had 45 genes as an input and 38 of them were shared among BRCA, PRAD and COAD datasets (see Appendix 1). In accordance, it can be argued that these three cancer types share a common change in their methylation pattern; hence this common effect of methylation should not be overlooked while investigating breast, prostate and colon cancers.

We have observed variety of regulatory genes when looked at all 45 differentially methylated genes. In order to investigate these genes in a more general way, we have conducted functional annotation of these genes searching for associated GO terms. When looked at the annotation results of genes with methylation change more than 25% (Table 4-8), we have detected cell differentiation and regulation of cell differentiation located at the top of the list, hence genes that were active in these pathways may be crucial for overall cancer disease mechanism.

GO annotation term	GO id	P-value
cell differentiation	GO:0030154	0.000234
regulation of transport	GO:0051049	0.000832
regulation of cell differentiation	GO:0045595	0.00099
positive regulation of transport	GO:0051050	0.002271
plasma membrane organization	GO:0007009	0.002493
digestive system development	GO:0055123	0.003816
muscle cell development	GO:0055001	0.00389
regulation of multicellular organismal development	GO:2000026	0.004487
regulation of cellular component organization	GO:0051128	0.00527
kinase activator activity	GO:0019209	0.006857
actin filament bundle	GO:0032432	0.007088
actomyosin	GO:0042641	0.008295
muscle tissue morphogenesis	GO:0060415	0.008547
heart growth	GO:0060419	0.008547
positive regulation of cellular process	GO:0048522	0.009531
actin binding	GO:0003779	0.009634

Table 4-8 Functional categorization results for the genes that are significantly altered for at least three types of cancers and have more than 25% methylation change. Analysis is performed using ConsensusPathDB "over-representation" analysis. GO: Biological Process annotation terms, ids and associated P-values are shown on the above table.

On the other hand, we have conducted an analysis for the commonly altered genes which were shared by at least 3 cancer types and which had significant expression change, such as log2 fold change >1 (Table 4-9). As a result, there were a total of 7 genes found shared by at least 3 cancer types and in difference to methylation results, 6 of these 7 genes were commonly altered in breast, thyroid and colon cancer datasets. Although the number of commonly altered genes was not as high as it was in methylation results, we can still argue that commonly altered six genes in breast, thyroid and colon cancers may be vital for cancer disease therapeutics of these cancer types.

GeneName	BRCA	THCA	COAD	PRAD	Gene Description
GPR115	3.482	4.823	1.435	-1.709	
MS4A15	3.630	4.828	2.871	-	May be involved in signal transduction
GDF15	2.970	-	2.080	1.930	Regulate tissue differentiation and maintenance
SERPINB2	1.322	2.399	1.731	-	Inhibits urokinase-type plasminogen activator
CDH11	1.858	1.654	1.684	-	Bone development and maintenance
TMEM105	1.371	2.028	2.639	-	Transmembrane Protein
IGFL2	2.558	5.464	4.371	-	Cellular energy metabolism. Growth. Development

Table 4-9 List of significantly expressed genes (FDR<0.05) for thyroid, breast, colon and prostate cancers. The values represent corresponding log fold changes. Only the genes that are shared by more than two cancer types are shown at this table. In addition, descriptions associated with each gene are also added to table using Genecards suite.

Most importantly, when the focus was on commonly altered pathways in our analysis (Table 4-7), erbB signalling pathway showed high significance and was found significantly affected for all cancer types in our dataset. When looked at the literature, erbB signalling pathway is reported to as active in various cancer types (Agus et al., 2002; Spano et al., 2005; Zhu, Zhu, Kim, Meltzer, & Cheng, 2014). Additionally, neurotrophin signalling pathway, focal adhesion and complement coagulation cascades are found as commonly affected for breast, thyroid and colon cancers, hence all of these pathways were reported as being involved in corresponding cancer mechanisms (Golubovskaya, Kweh, & Cance, 2009; Molloy, Read, & Gorman, 2011). Similarly, ECM-receptor interaction and arrhythmogenic right ventricular cardiomyopathy were found as commonly altered for breast, colon and prostate cancers. Interestingly, we have found non-small cell lung cancer and prostate cancer pathways as common for breast, thyroid and prostate cancers, implying all these cancers might be sharing similar mechanisms. Chronic myeloid leukaemia, glioma, T cell receptor signalling pathway and bacterial invasion of epithelial cells are also found as altered for these cancer types.

Furthermore, G protein-coupled receptors are shown as being involved in cancer (Dorsam & Gutkind, 2007), which are known for their tumour suppressor and oncogenic roles (Feigin, 2013). At our analysis, we have identified GPR115 as significantly altered and having methylation change >25%, which implies GPR115 may be crucial for the investigated cancer types. In more detail, GPR115 expression was down-regulated for three cancer types whereas it was up-regulated in prostate cancer preserving the inverse correlation between

methylation and expression. Since GPCR proteins can act as both oncogene and tumour suppressor, inspective research on the role of GPR115 gene can provide deeper insights into the cancer biology. In order to validate the significance of GPR115 gene, we have also questioned the mutational frequency information around GPR115 gene. Located at the 6^{th} chromosome between 47,685,864 and 47,722,021 base pairs, there were only five variations found in dbSNP database, which were all located in introns. There was not any known damaging variation (indels, start-stop mutations, frameshift mutations) observed in GPR115 gene, which supports our findings that GPR115 may be crucial for cancer disease etiology.

5. IDENTIFYING THE METHYLATION-DRIVEN PATTERNS IN CANCER

5.1 Results

At this analysis, we have set Benjamini-Hochberg False Discovery Rate <0.1 as "differential methylation" threshold not being too stringent, and disregarding the expression significances, we have only used the genes having expression fold change >2. The reason behind was to observe the major effect of methylation to the expression levels.

5.1.1 Differential Expression & Methylation Analysis

We identified the significantly differentially methylated and differentially expressed genes in each cancer type separately. The numbers of differentially methylated and differentially expressed genes are provided in Table 5-1.

Because we desired to investigate the interplay between methylation and expression, in further analysis, we continued with the genes that showed both significant methylation and expression. More specifically, renal papillary cell carcinoma exhibited the greatest number of differentially methylated genes (FDR<0.1), while the number of differentially expressed genes was highest for cholangiocarcinoma, and the thyroid cancer dataset contained the minimum numbers of genes regarding both differential expression and differential methylation.

Cancer Type	Number of differentially methylated Genes	Number of differentially expressed genes	Genes with both differential expression and methylation
Cholangiocarcinoma	16,796	9207	8050
Colon adenocarcinoma	15,761	5657	4789
Liver Hepatocellular carcinoma	17,206	5252	4598
Lung squamous cell carcinoma	15,642	8403	6940
Renal papillary cell carcinoma	17,558	5785	5072
Thyroid cancer	14,825	3616	2951

Table 5-1 Numbers of differentially methylated and differentially expressed genes for each cancer type. The rightmost column provides information about the numbers of genes that were both differentially expressed and differentially methylated.

We set a methylation change of 32.2% as the large methylation change threshold, and we set a 15% methylation change as the normal methylation change threshold. The latter value was previously defined at our first work. The numbers of genes showing methylation changes exceeding 15% and 32.2% for each cancer type are illustrated in Table 5-2.

Cancer Type	Number of genes with a >32.2% methylation change	Number of genes with a >15% methylation change
Cholangiocarcinoma	1792	4920
Colon adenocarcinoma	1503	3773
Liver hepatocellular carcinoma	852	3375
Lung squamous cell carcinoma	1770	5490
Renal papillary cell carcinoma	524	3054
Thyroid cancer	173	1001

Table 5-2 Numbers of genes showing methylation changes exceeding 32.2% and 15%. Only the genes exhibiting both differential methylation and differential expression were included in this analysis.

The results revealed that, in terms of the numbers of genes showing more than a 15% methylation change, colon adenocarcinoma exhibited the greatest percentage of genes exhibiting a large methylation change (23%), whereas for renal papillary cell carcinoma and thyroid cancer, the percentage of large methylation changes was below 8%.

5.1.2 Distances to Drivers

Regardless of the direction of the change, it is well established that differences in methylation levels between control and tumor samples exert an important influence on expression levels. However, whether large methylation changes elicit more direct effects and interact more strongly with driver genes remains unknown. To address these issues, we adopted a graph-based approach and calculated the distances between large methylation changes and potentially cancer-causing driver genes by considering pairwise protein-protein interactions. Moreover, we compared the effects of large methylation changes (>32.2%) and normal methylation changes (>15%, <32.2%); the corresponding results are provided in

Table 5-3. In this analysis, we included only the genes that exhibited an inverse correlation between expression and methylation.

Cancer Type	32.2% methylation change; genes with a path to driver <8 vs. path to driver <4	Average Distance from driver genes (distance<3); >32.2% methylation change	15% methylation change; genes with a path to driver <8 vs. path to driver <4	Average Distance from driver genes (distance <3); >15% methylation change	Difference in average distance between 32.2% and 15%
CHOL	126/1119=0.11	2.07	215/3802=0.05	2.15	-0.08
COAD	51/796 =0.06	2.05	51/2978=0.02	2.10	-0.05
LIHC	21/571=0.04	1.76	58/2854=0.02	2.00	-0.24
LUSC	108/923=0.12	2.00	179/4559=0.04	2.06	-0.06
KIRP	24/371=0.06	1.88	65/2754=0.02	2.12	-0.24
THCA	12/148=0.08	2.58	27/854=0.03	2.21	+0.37

Table 5-3 Numbers of genes with large methylation changes (32.2%) and normal methylation change (15%) that reached the driver genes and the average distances between them.

The results revealed that the genes with large methylation changes affected driver genes in fewer steps, except for thyroid cancer. Specifically, the average distance from potentially cancer-causing genes was found to be below 2 for liver hepatocellular carcinoma and renal cell papillary carcinoma. When the focus was placed on the proportion of genes that interacted with driver genes in fewer than 4 steps, we observed an overall tendency for more direct interactions when large methylation changes were present. For the genes with normal methylation changes, the proportion of close interactions was 3.00%, while for genes exhibiting large methylation changes the average proportion was 7.83%.

Moreover, with the aim of testing the significance of our findings, we randomly selected differentially altered genes from each cancer type. Subsequently, we examined the average distances from driver genes and the proportions of close interactions with driver genes by applying the same procedure 100 times. During the random selection process, we addressed each cancer type separately, and to determine the number of randomly selected genes, we considered all genes with large methylation changes and a path to a driver of <4, as in our

original analysis (numbers of genes for each cancer type: CHOL: 140, COAD: 53, LISC: 23, LUSC: 123, KIRP: 27, and THCA: 13). Consequently, despite repeating the same procedure 100 times with different genes, only single paths to the drivers were found for the LUSC and CHOL datasets. More interestingly, for the other four datasets, there were no paths to driver genes. Thus, these results support our findings, and overall, we can state that the genes with large methylation changes interacted with driver genes in higher proportions and in a more direct manner. Thus, large methylation changes are crucial to cancer etiologies.

5.1.3 Affected Suppressors and Oncogenes

To provide insights into cancer mechanisms, we focused on tumor suppressors with decreasing expression levels and oncogenes with increasing expression levels. The tumor suppressors showing expression fold-changes greater than -2 and the oncogenes with expression fold-changes greater than 2 and with distances from a driver gene to a large methylation change-driven gene including a maximum of 3 steps are illustrated in Table 5-4. Only the genes that were shared by at least two cancer types are presented in this table.

	THCA	**CHOL**	**COAD**	**KIRP**	**LUSC**	**Liver**
AGTR1	3	2	2	-	2	1
IGF1	-	0	1	1	1	1
CXCL12	-	1	1	-	1	2
FGFR3	2	0	-	-	0	-
EPAS1	-	0	0	-	0	-
SRC	-	0	-	0	-	1
FOXD3	-	-	0	0	-	0
PPARG	1	1	-	-	-	-
FOXO1	-	1	-	0	-	-
CDK4	-	1	-	-	2	-
NKX3-1	-	1	-	-	-	1
PIK3R1	-	0	-	-	0	-
PRKCB	-	-	1	-	1	-
EDNRB	-	-	0	-	0	-

Table 5-4 Numbers of steps between driver genes and genes with large methylation changes for genes that were shared by at least two types of cancer. Moreover, the genes that were shared and the cancer types sharing these genes can be extracted from this table.

Moreover, the identities of methylation-driven genes, the pathways they affected, and the common mechanisms among different cancer types are illustrated according to color in a

general illustration of the KEGG: Pathways in Cancer (Figure 5-1). Additionally, the genes that were shared, the cancer types among which they were shared, and cancer-specific genes are illustrated in Figure 5-2. Thus, we observed that the AGTR1 (GPCR protein) and IGF1 genes were shared by 5 different cancer types, which are shown in red. CXCL12 was identified in the CHOL, COAD, LUSC and LIHC datasets and is shown in orchid. The genes that were shared by three different cancers (i.e., FGFR3, EPAS1, SRC, and FOXD3) are shown in coral, and the genes that were shared by two types of cancers (i.e., PPARG, FOXO1, CDK4, NKX3-1, PIK3R1, PRKCB and EDNRB) are shown in light pink.

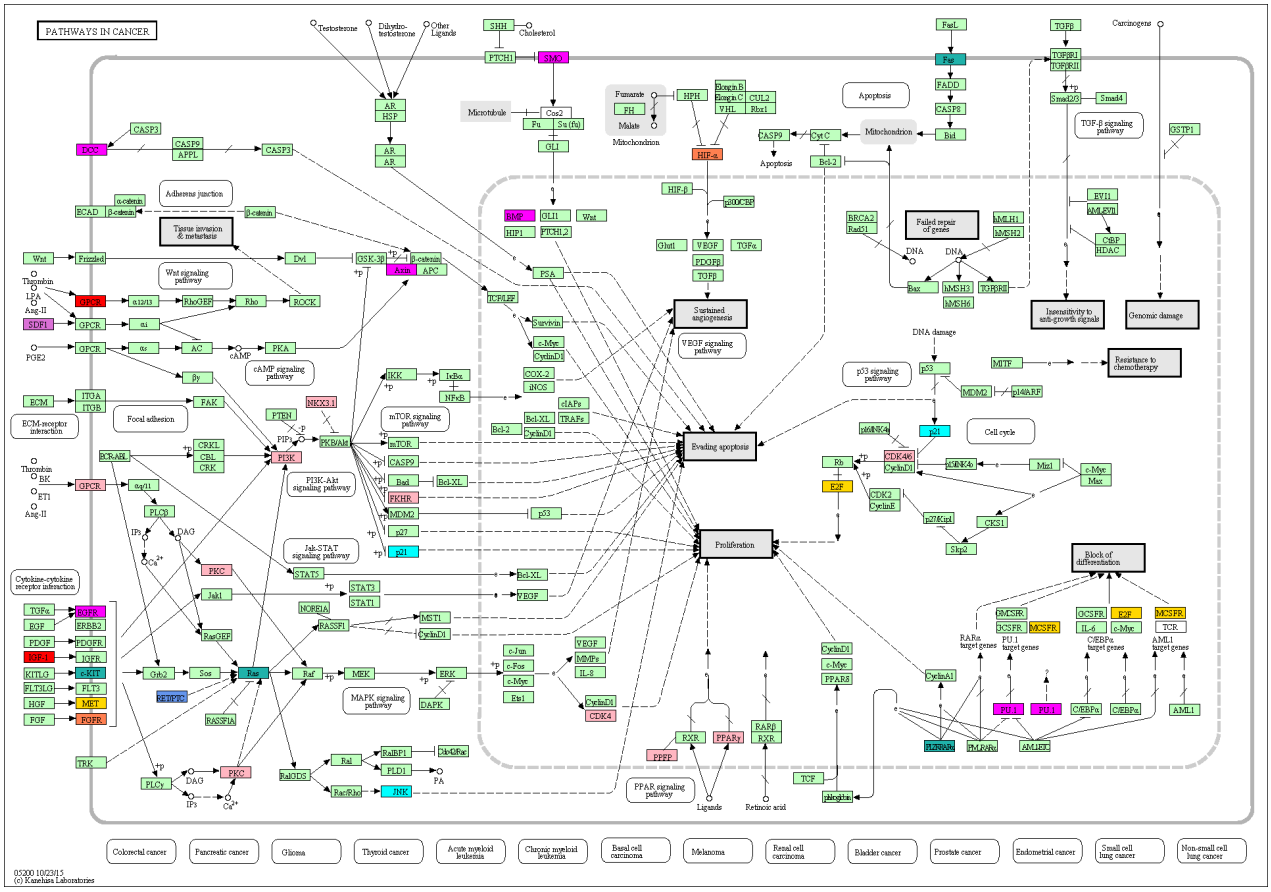

Figure 5-1 KEGG: Pathways in Cancer: A general picture. The genes are color-coded according to the number or cancer types among which they are shared. Red indicates sharing by 5 cancer types; orchid indicates sharing by 4 cancer types; coral indicates sharing by 3 cancer types; and light pink indicates sharing by 2 cancer types. In contrast, the genes that were affected only in a single cancer type are represented with the following colors: only THCA, cornflower blue; only CHOL, light sea green; only COAD, cyan; only KIRP, gold; and only LUSC, magenta. Unfortunately, there were no genes that were specific to LIHC.

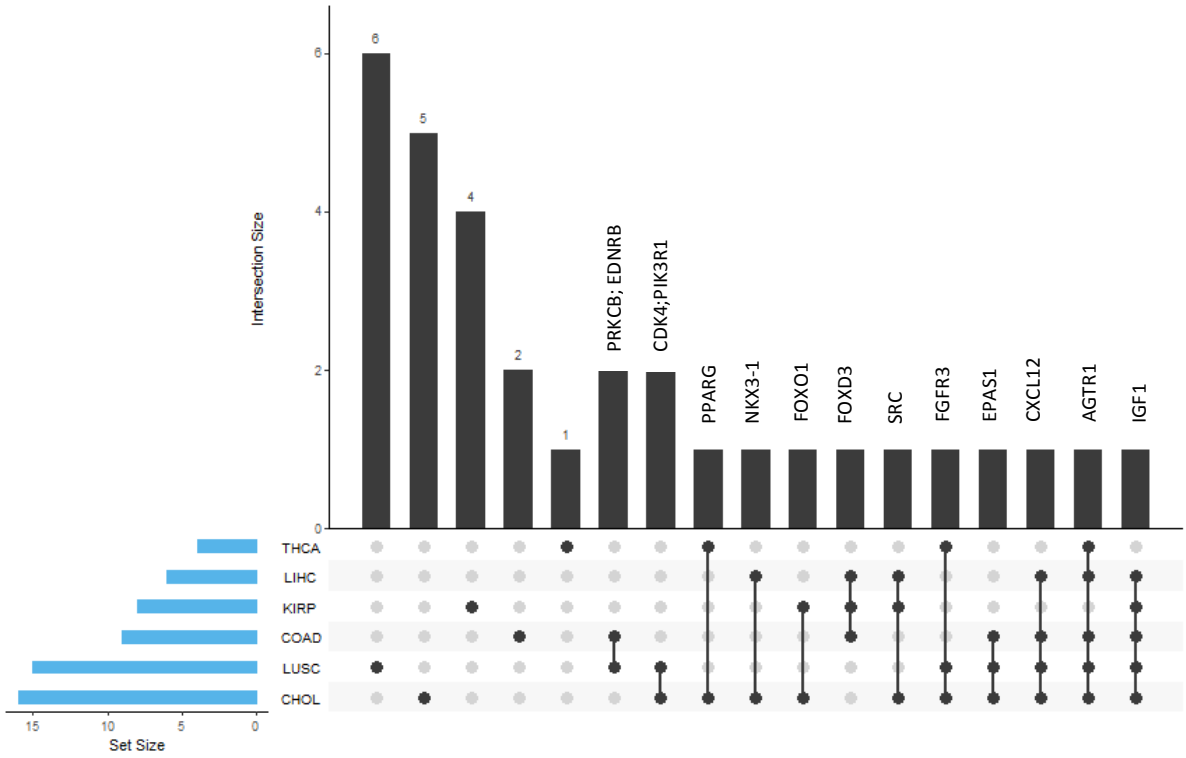

Figure 5-2 Diagram illustrating the genes shared by different cancers and the cancers among which they are shared.

Focusing on the affected genes that were observed only in a single cancer type revealed that the RET gene was affected by high methylation in thyroid cancer (shown in cornflower blue). More specifically, the RET gene interacts with the Ras gene, which is involved in various cancer mechanisms. Examining cholangiocarcinoma, we observed that the HRAS, KIT, ZBTB16, FAS and NCOA4 genes were altered by high methylation (shown in light sea green). Similar to thyroid cancer, the Ras gene was also found to be affected in cholangiocarcinoma. The ZBTB16 gene is active in the blocking of differentiation; thus, methylation-driven abnormalities in ZBTB16 may lead to cholangiocarcinoma. We observed that MAPK10 and CDKN1A were only altered in colon cancer (shown in cyan). Similar to cholangiocarcinoma, four genes that were detected only in renal cell carcinoma play crucial roles in the blocking of differentiation and proliferation (shown in gold). In lung squamous cell carcinoma, we detected 15 genes that were altered by high methylation, and six of these genes were observed to be affected only in this type of cancer (shown in magenta). In contrast, in the hepatocellular carcinoma dataset, we did not detect any gene that was only affected in liver cancer.

Moreover, we conducted copy number analyses of the driver genes that we detected. Because increases in copy number are associated with increases in expression, we asked whether the main reason for the observed expression changes was copy number aberrations, rather than large methylation changes. Copy number analyses corresponding to each driver gene are provided in Appendix 2. We focused on the genes showing copy number alterations in more than half of the datasets. Only in the lung cancer dataset were we able to detect such genes. More specifically, the tumor suppressor AGTR1 was found as increased in 5 of the 7 tumor samples, while the tumor suppressor SPI1 was found as increased in 4 samples. In contrast, there were 3 samples showing decreased copy numbers of the EPAS1 gene; hence, decreases in the expression of this gene may be related to decreases in copy numbers. Additionally, in the thyroid cancer dataset, copy number analysis of the RET gene revealed decreases in 5 of the 46 samples, while no sample exhibited an increase in the RET copy number. These findings imply that the expression of this gene should be downregulated. However, using our graph-based approach, we were able to associate an increase in RET gene expression with a large decrease in methylation in the proximity of the RET gene.

5.1.4 Cancer Similarity Analysis

Although the affected suppressors and oncogenes provide a general picture regarding shared cancer mechanisms, we conducted cluster analysis to focus on the similarities between the different cancer types. The results of the correlation analyses of differential expression, methylation and PPIs are illustrated in Figure 5-3. Examination of the clustering of differential expression revealed that cholangiocarcinoma (CHOL) and liver hepatocellular carcinoma (LIHC) shared the greatest percentage of expression patterns, with 60%. Additionally, thyroid cancer (THCA) and colon adenocarcinoma (COAD) clustered together with a 42% correlation. In contrast, renal cell papillary carcinoma (KIRP) and lung squamous cell carcinoma (LUSC) were separated from the other cancer types when only the differential expression values were examined.

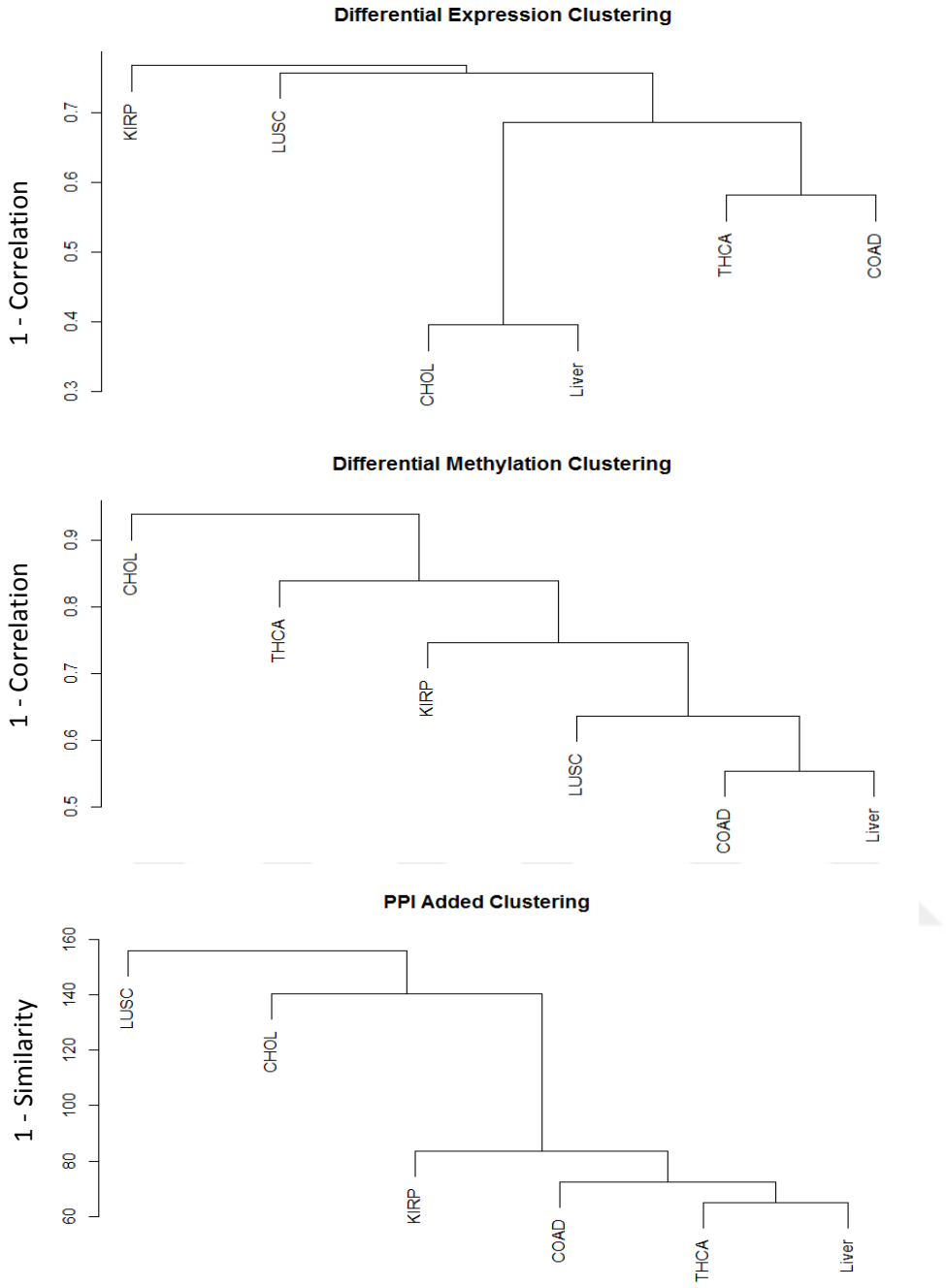

Figure 5-3 Clustering results for differential expression, differential methylation and PPI included analysis

Interestingly, the clustering of differential methylation was not similar to the clustering of differential expression. Colon adenocarcinoma and liver hepatocellular carcinoma exhibited the highest correlation, with 45% similarity, which implies that these cancers may be associated with similar methylation changes. There were no other significant

correlations between pairs of cancer types; thus, LUSC, KIRP, THCA and CHOL were individually added to the first cluster in that order.

Finally, clustering analyses according to the integration of the protein-protein interaction networks and methylation and expression information revealed that THCA, LISC, COAD and KIRP were closest to each other, which implies that the patterns that lead to disease states in these cancers are similar. In contrast, CHOL and LUSC were separated from these cancers; thus, the methylation-driven gene sets that lead to these cancer states follow distinct patterns.

Additionally, we calculated the correlations between the different cancer types using 110 randomly selected genes after executing the same procedures 1000 times. These 1000 results were then averaged and compared with our original results. As illustrated in Table 5-5, the overall correlation of the differential expression results decreased, whereas the correlation of the differential methylation results increased.

	DIFFERENCES IN EXPRESSION CORRELATIONS BETWEEN THE NORMAL AND 1000 RANDOM EXECUTION ANALYSES					
	THCA	CHOL	COAD	KIRP	LUSC	Liver
THCA	0	0.064	0.182	0.099	-0.075	0.294
CHOL		0	0.057	0.053	0.082	0.053
COAD			0	0.006	0.109	0.074
KIRP				0	0.052	0.029
LUSC					0	-0.187
LIVER						0
	DIFFERENCES IN METHYLATION CORRELATIONS BETWEEN THE NORMAL AND 1000 RANDOM EXECUTION ANALYSES					
	THCA	CHOL	COAD	KIRP	LUSC	Liver
THCA	0	-0.205	-0.066	-0.030	-0.190	-0.042
CHOL		0	-0.108	-0.142	-0.033	-0.077
COAD			0	0.035	-0.048	-0.020
KIRP				0	0.053	-0.034
LUSC					0	-0.044
LIVER						0

Table 5-5 Differences in correlations between the original work and the 1000 random execution analyses shown for each pairwise relationship.

5.2 Discussion & Conclusion

Our aim in the present study was to identify methylation-driven mechanisms in cancer by adopting a network-based approach. Moreover, we searched for similarities and differences between different cancer types by focusing on the mechanisms affecting cancer-related genes.

In examining the effects of methylation on cancer, driver genes that are crucial for cancer progression should be defined prior to the analysis. Although a variety of driver prediction algorithms are available in the literature, most are based on predicting the effects of mutations, and there is no consensus regarding methylation-based drivers. At this point, the use and annotation of cancer-associated genes depending on their suppressor or oncogene status may provide deeper insight into disease etiology compared with mutation-based prediction algorithms. For this purpose, we extracted a set of genes that had previously been associated with cancer from the KEGG: Pathways in Cancer dataset and through literature mining.

Second, we examined data regarding the differential expression and methylation of these cancer-related genes in six different cancer datasets. Prior to further analysis, we integrated SNP information into the analysis and excluded CpG regions showing mutation frequencies greater than 0.1 in the 1000 Genomes database. During the extraction of differentially methylated and differentially expressed genes, setting an FDR threshold for methylation and fold-change threshold for expression did not decrease the number of short listed genes as desired; therefore, a further decrease in the number of genes was necessary (Table 5-1). Given this information, setting a methylation change threshold appeared to be promising means of obtaining a clearer picture of the genes that are altered in cancer. Additionally, based on the idea that large methylation changes might exhibit more rapid effects on the mechanisms leading to cancer states, we defined a large methylation change threshold. To this end, we calculated the central value and scaling parameter by pooling all of the data. Consequently, we arrived at a central value of 16.8% and a scaling parameter of 7.77%. Similar to the normal distribution, we set the central value plus two scaling parameters (i.e., a 32.2% methylation change) as the high methylation threshold.

Subsequently, we continued our analysis of the effects of methylation-driven changes on cancer-related genes (driver genes). In this analysis, when no methylation-driven changes in gene expression were identified in the driver genes themselves, we investigated whether

integrating protein-protein interaction information could provide additional information about the underlying trigger mechanisms leading to significant expression changes in cancer-related genes. Applying this approach, we used the widely accepted STRING database and calculated the numbers of steps required for genes with methylation changes greater than 32.2% to reach driver genes. To compare large and normal methylation changes, we applied the same procedure utilizing a 15% methylation change threshold. We found that large methylation changes exerted a major influence on the expression of driver genes in fewer steps (1.95 vs. 2.09, except for thyroid cancer) and at a higher proportion (7.83% vs. 3.00%), although these results varied among different cancer types. Moreover, to test the significance of our findings, we applied the same procedure following the random selection of genes 100 times. As a result of this analysis, we found that the ratio of selected genes that reached driver genes in fewer than 3 steps was almost 0 for all of the examined cancer types.

To validate our findings regarding the identified tumor suppressors and oncogenes, we searched for support in the literature. Using our novel graph-based approach, we observed a decrease in the expression level of the tumor suppressor insulin-like growth factor 1 (IGF1), and this decrease was primarily driven by a large methylation change in 5 different types of cancer. Regarding the other cancer types without methylation-driven mechanisms related to IGF1, the expression level of IGF1 was identified as -0.87. Because 0.87 was below our expression status threshold (i.e., a log 2 fold-change > 1), this gene was excluded from our analysis in the beginning. When we examined the literature, we found that IGF1 has previously been associated with proliferation and apoptosis, and it has also been demonstrated to play a crucial role in cholangiocarcinoma (Alvaro et al., 2006), colon cancer (Giovannucci, 2001), kidney cancer (Major, Pollak, Snyder, Virtamo, & Albanes, 2010), lung squamous cell carcinoma (Heist, Sequist, & Engelman, 2012) and liver hepatocellular carcinoma (Stuver et al., 2000). Similarly, angiotensin II receptor 1 (AGTR1) was identified by our method as being shared by 5 different cancer types. AGTR1 is primarily involved in the renin-angiotensin system and has previously been validated as an important tumor suppressor in lung cancer (Guo et al., 2015), cholangiocarcinoma (Okamoto et al., 2010), colon cancer (Dai et al., 2015) and liver cancer (Yoshiji et al., 2011). In the present study, AGTR1 was found to exhibit a decreased expression level; hence, in a maximum of three steps, we identified genes that were altered by large methylation changes that signaled decreases in the expression level of AGTR1. The only cancer that lacked a methylation-driven pattern for AGTR1 was kidney cancer, although the expression level of AGTR1 in

kidney cancer was identified as -3.8 and was probably influenced by tissue-specific alterations. In summary, our findings suggest that both AGTR1 and IGF1 can be used as indicators of overall cancer progression.

Additionally, it has been previously demonstrated that CXCL12 expression suppresses pancreatic cancer growth and metastasis (Roy et al., 2014). In our analysis, the expression of the potential tumor suppressor chemokine ligand 12 (CXCL12) was found to be decreased, possibly due to high methylation, in four different cancer types. In the other two cancer types (particularly thyroid cancer) the CXCL12 gene was not included in the analysis because the change in the expression level of CXCL12 was slightly below our threshold (-0.0996). Regarding the kidney cancer data, the expression of CXCL12 was found to be significantly decreased (-1.15); however, the underlying cause of this alteration was not high methylation, but rather, normal methylation (0.18). Moreover, increased methylation-driven expression of the important oncogene SRC was identified in cholangiocarcinoma, hepatocellular carcinoma and colon adenocarcinoma, and SRC has previously been demonstrated to be active in cancer progression (Irby & Yeatman, 2000).

Among the individual, cancer-specific genes that were detected in our graph-based analysis, the well-known RET oncogene has been associated with thyroid cancer progression in numerous studies (A, F, G, & M, 2011; Kimura et al., 2003). Regarding cholangiocarcinoma, HRAS, KIT and FAS have previously been associated with cholangiocarcinoma in the literature (Jhala, Vickers, Argani, & McDonald, 2005; Mansuroglu et al., 2009; Tada, Omata, & Ohto, 1992). Additionally, the most recent efforts directed toward the CDKN1A gene have demonstrated that downregulation of this gene leads to colon cancer progression (Kramer et al., 2016), and similar results have been reported for MAPK10 (J. Y. Fang & Richardson, 2005). Both of these genes were found to be differentially altered by high methylation in colon cancer in our analysis. The C-met oncogene encodes the hepatocyte growth factor receptor (Linehan, Srinivasan, & Schmidt, 2010), and we observed increased MET levels in the kidney cancer data. Additionally, WT1 interacts with the p53 gene, and over-expression of WT1 is associated with renal cell carcinoma (Campbell et al., 1998). The CSF1R gene has previously been studied and linked to renal cell carcinoma as a potential therapeutic target in patients (Soares et al., 2009). Similarly, the E2F1 gene has been suggested as a potential therapeutic target and highlighted as playing a key role in renal cell carcinoma (Ma et al., 2013). Although a total of six genes were found to be differentially altered in our liver cancer data, none of these genes were

found to be liver cancer-specific in our analysis. More specifically, 5 of the 6 genes were found to be shared with cholangiocarcinoma, which implies that these two cancers share similar mechanisms of cancer progression. In contrast, we identified six cancer-specific genes for lung squamous cell carcinoma, and we have found support in the literature for each of these genes. More specifically, downregulation of the tumor suppressor BMP4 has been linked to LUSC; thus, our method also explained the underlying methylation-driven mechanism (W. T. Fang et al., 2014). Additionally, changes in EGFR have been associated with lung cancer, and this gene has been recommended as a drug target (Paez et al., 2004). Alterations of AXIN2 have been demonstrated to contribute to carcinogenesis, specifically in lung cancer (Gunes, Pinarbasi, Pinarbasi, & Silig, 2009), and the downregulation of this suppressor via a large methylation change was successfully detected by our method.

Moreover, we investigated whether increases and decreases in expression were caused by copy number aberrations, rather than large methylation changes. We focused on the driver genes that were identified as crucial to cancer etiology via our graph-based strategy. Because copy number changes in single samples are generally neglected due to corresponding statistical weakness, we focused on genes showing copy number changes in more than half of the datasets. Most notably, we observed genes that met the aforementioned conditions only in lung cancer. More specifically, the AGTR1 and SPI1 genes were identified as showing copy number increases in 5 and 4 samples, respectively, whereas the expression of these tumor suppressors was downregulated in the LUSC dataset. Furthermore, we associated these changes with large methylation increases. Because an increase in copy number is generally associated with greater expression, our findings suggest that large methylation changes were the predominant factors underlying the downregulation of these tumor suppressors, which was supported by copy number analyses. Remarkably, a similar situation was observed for the RET gene in the thyroid cancer dataset; the expression of the RET oncogene was upregulated, whereas 5 tumor samples exhibited decreased copy numbers of RET. Using our approach, we were able to associate the increase in RET oncogene expression with a large methylation decrease, contributing to an improved understanding of the etiology of thyroid cancer.

In contrast to previously applied methods, such as MethylMix and LRpath, our approach is the only strategy that has integrated protein-protein interaction information in the identification of methylation-driven genes. Recently, many studies have benefited from the use of DNA methylation changes as prognostic markers of disease progression (Bartlett et al.,

2015; De et al., 2013; Kanemoto et al., 2014; Marcucci et al., 2014; Wei et al., 2015). Examples include studies of B cell lymphoma, acute myeloid leukemia, glioblastoma and epithelial squamous cell carcinoma. Basically, Cox regression models that benefit from a set of marker genes were applied in these studies to predict survival and cancer stage. A similar approach can be applied to the genes we have identified, and this method of predicting survival and disease stage can be extended to hepatocellular carcinoma, lung squamous cell carcinoma, colon adenocarcinoma, cholangiocarcinoma and thyroid cancer. Moreover, oxidative stress, which has tight links with dietary and environmental factors, has been previously associated with changes in DNA methylation level (Campos et al., 2007; Franco, Schoneveld, Georgakilas, & Panayiotidis, 2008; Khor et al., 2014; Wongpaiboonwattana, Tosukhowong, Dissayabutra, Mutirangura, & Boonla, 2013). Thus, one can argue that changing the dietary habits may be helpful at preventing the methylation-caused cancers. At this point, further research is required for building associations between diets and cancer prevention. Especially the genes that we have extracted as important using our graph-based method may be beneficial while tracking the effects of different diets on epigenome.

The goal of the last stage of our analysis was the identification of similarities and differences between the different cancer types, with a focus on driver genes that were classified as either “tumor suppressors and downregulated” or “oncogenes and upregulated”. We explored whether the same driver genes were active in different cancer types and whether the same sets of interactions were triggered by large methylation changes, leading to a cancer state. When the results of the analyses of differential expression, differential methylation and PPI-included clustering were examined, we found that liver, colon, kidney and thyroid cancer were similar in the results from the analysis that included the PPIs, which implies that similar mechanisms are involved in these cancers and may lead to cancer progression. Additionally, these cancers were not far from each other in terms of the differential expression and differential methylation results. In contrast, although lung cancer was observed to be close to other cancers in the differential methylation results, the situation was not the same for the differential expression results. Additionally, the situation was reversed in the comparison of cholangiocarcinoma with LUSC. CHOL was found to be close to other cancers in the differential expression results; however, in the differential methylation results, CHOL was one of the most distant cancers. Examination in greater detail revealed that cholangiocarcinoma was especially close to liver cancer in the differential expression results; however, because the methylation profiles were not similar, the similarity between these

cancers was increased in the PPI-included results. Overall, it can be argued that because LUSC and CHOL were found to be isolated from the other four cancers, tissue-specific factors may play major roles in determining the expression of genes in these cancers. Additionally, greater numbers of genes with large methylation changes were identified in these cancer types; therefore, alternative methylation-driven paths to driver genes were identified in these cancer types and led to divergence. In contrast, there were fewer genes with large methylation changes among the other cancer types, and these genes were also shared by different cancer types. Thus, we observed that these cancer types were closer to each other in the cluster analysis, and we argue that similar cancer mechanisms are active in these cancer types.

Overall, the synopsis of our findings suggests that in a list of cancer-related genes, methylation-driven pathways either affect the gene itself in a manner that promotes cancer development, or the contributions of the genes to cancer can be explained by cascades of methylation-driven events involving small numbers of interactions that trigger corresponding changes in cancer-related genes. The genes involved in different types of cancer vary, but the manner in which methylation-driven mechanisms affect driver genes exhibits similarities across all cancer types. Our graph-based, integrative approach for identifying methylation-driven patterns provides valuable information regarding cancer etiology, and the genes that are highlighted by our method (especially the oncogenes) may be used as potential therapeutic targets. Moreover, when cluster analysis was applied following the integration of protein-protein interaction information, colon, kidney, liver and thyroid cancers were found to be clustered together with high similarity; therefore, these cancers may share methylation-driven mechanisms that lead to a cancer state.

6. CONCLUSION, TAKE HOME LESSONS

In this study, in order to investigate the role of methylation in cancer mechanisms, we have examined the interplay between methylation and expression occurring at the genes themselves as well as the methylation changes occurring at neighbouring genes by considering protein-protein interaction information. First of all, we have revealed that in methylation-based experiments, introducing a methylation level change threshold leads to improved results by decreasing false positive calls. Secondly, we have observed significant methylation changes in wide variety of cancers. Moreover, we have shown that methylations occurring at 1st exon, TSS200 and 5'UTR regions show higher inverse correlation with expression, implying methylation change at these regions have stronger impact on expression. Most importantly, by adopting a novel graph-based approach, and investigating protein-protein interactions, we were able to explain most of the cancers by high methylation changes, which is crucial since DNA methylation is a reversible phenomenon. As a matter of fact, large methylation changes are of higher importance since they affect the cancer-causing genes more efficiently and more directly, which makes these genes promising candidates for cancer therapy. Interestingly in functional enrichment results of differentially methylated genes, post translational modifications and transcription factors are found significantly enriched, thus, methylation change may be rapidly affecting other genes by affecting the transcription factors. It is important to note that, we have encountered significant methylation-based changes especially in GPCR proteins in wide variety of cancers, namely ADGRF4 and AGTR1 genes, thus these genes may be crucial for cancer disease etiology. Overall, we propose a novel data integration and data analysis method involving graph-based analysis and functional enrichment. It is now shown in this study that, our findings regarding methylation-driven mechanisms in wide variety of cancers improves the knowledge of cancer disease etiology, which may be crucial for potential cancer therapeutics.

7. BIBLIOGRAPHY

A, T., F, S., G, P., & M, B. (2011). Genetic alterations in medullary thyroid cancer: diagnostic and prognostic markers. *Curr Genomics, 12*(8), 618-625. doi: 10.2174/138920211798120835

Agrez, M. V. (1996). Cell adhesion molecules and colon cancer. *Aust N Z J Surg, 66*(12), 791-798.

Agus, D. B., Akita, R. W., Fox, W. D., Lewis, G. D., Higgins, B., Pisacane, P. I., . . . Sliwkowski, M. X. (2002). Targeting ligand-activated ErbB2 signaling inhibits breast and prostate tumor growth. *Cancer Cell, 2*(2), 127-137.

Akiyama, Y., Maesawa, C., Ogasawara, S., Terashima, M., & Masuda, T. (2003). Cell-type-specific repression of the maspin gene is disrupted frequently by demethylation at the promoter region in gastric intestinal metaplasia and cancer cells. *Am J Pathol, 163*(5), 1911-1919. doi: 10.1016/S0002-9440(10)63549-3

Alashwal, H., Dosunmu, R., & Zawia, N. H. (2012). Integration of genome-wide expression and methylation data: relevance to aging and Alzheimer's disease. *Neurotoxicology, 33*(6), 1450-1453. doi: 10.1016/j.neuro.2012.06.008

Almond, J. B., & Cohen, G. M. (2002). The proteasome: a novel target for cancer chemotherapy. *Leukemia, 16*(4), 433-443. doi: 10.1038/sj.leu.2402417

Alvaro, D., Barbaro, B., Franchitto, A., Onori, P., Glaser, S. S., Alpini, G., . . . Gaudio, E. (2006). Estrogens and insulin-like growth factor 1 modulate neoplastic cell growth in human cholangiocarcinoma. *Am J Pathol, 169*(3), 877-888. doi: 10.2353/ajpath.2006.050464

Ames, E., Hallett, W. H., & Murphy, W. J. (2009). Sensitization of human breast cancer cells to natural killer cell-mediated cytotoxicity by proteasome inhibition. *Clin Exp Immunol, 155*(3), 504-513. doi: 10.1111/j.1365-2249.2008.03818.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet, 25*(1), 25-29. doi: 10.1038/75556

Badal, V., Chuang, L. S., Tan, E. H., Badal, S., Villa, L. L., Wheeler, C. M., . . . Bernard, H. U. (2003). CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression. *J Virol, 77*(11), 6227-6234.

Bakir-Gungor, B., Egemen, E., & Sezerman, O. U. (2014). PANOGA: a web server for identification of SNP-targeted pathways from genome-wide association study data. *Bioinformatics, 30*(9), 1287-1289. doi: 10.1093/bioinformatics/btt743

Barach, Y. S., Lee, J. S., & Zang, X. (2011). T cell coinhibition in prostate cancer: new immune evasion pathways and emerging therapeutics. *Trends Mol Med, 17*(1), 47-55. doi: 10.1016/j.molmed.2010.09.006

Bartlett, T. E., Jones, A., Goode, E. L., Fridley, B. L., Cunningham, J. M., Berns, E. M., . . . Widschwendter, M. (2015). Intra-Gene DNA Methylation Variability Is a Clinically Independent Prognostic Marker in Women's Cancers. *PLoS One, 10*(12), e0143178. doi: 10.1371/journal.pone.0143178

Belinsky, S. A., Klinge, D. M., Stidley, C. A., Issa, J. P., Herman, J. G., March, T. H., & Baylin, S. B. (2003). Inhibition of DNA methylation and histone deacetylation prevents murine lung cancer. *Cancer Res, 63*(21), 7089-7093.

Benjamini Y, H. Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society B., 57*, 289-300.

Bloch, D. (1966). A Note on the Estimation of the Location Parameter of the Cauchy Distribution. *Journal of the American Statistical Association, 61*(315), 852-855. doi: 10.2307/2282794

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., & Scandura, J. M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One, 6*(1), e14524. doi: 10.1371/journal.pone.0014524

Britten, C. D. (2004). Targeting ErbB receptor signaling: a pan-ErbB approach to cancer. *Mol Cancer Ther, 3*(10), 1335-1342.

Butcher, L. M., & Beck, S. (2015). Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods, 72*, 21-28. doi: 10.1016/j.ymeth.2014.10.036

Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G., & Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet, 21*(1), 103-107. doi: 10.1038/5047

Campbell, C. E., Kuriyan, N. P., Rackley, R. R., Caulfield, M. J., Tubbs, R., Finke, J., & Williams, B. R. (1998). Constitutive expression of the Wilms tumor suppressor gene (WT1) in renal cell carcinoma. *Int J Cancer, 78*(2), 182-188.

Campos, A. C., Molognoni, F., Melo, F. H., Galdieri, L. C., Carneiro, C. R., D'Almeida, V., . . . Jasiulionis, M. G. (2007). Oxidative stress modulates DNA methylation during melanocyte anchorage blockade associated with malignant transformation. *Neoplasia, 9*(12), 1111-1121.

Cancer Genome Atlas, N. (2012a). Comprehensive molecular characterization of human colon and rectal cancer. *Nature, 487*(7407), 330-337. doi: 10.1038/nature11252

Cancer Genome Atlas, N. (2012b). Comprehensive molecular portraits of human breast tumours. *Nature, 490*(7418), 61-70. doi: 10.1038/nature11412

Cancer Genome Atlas Research, N. (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature, 489*(7417), 519-525. doi: 10.1038/nature11404

Cancer Genome Atlas Research, N. (2014). Integrated genomic characterization of papillary thyroid carcinoma. *Cell, 159*(3), 676-690. doi: 10.1016/j.cell.2014.09.050

Cancer Genome Atlas Research, N., Linehan, W. M., Spellman, P. T., Ricketts, C. J., Creighton, C. J., Fei, S. S., . . . Zuna, R. (2016). Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med, 374*(2), 135-145. doi: 10.1056/NEJMoa1505917

Chang, H. J., Yoo, B. C., Lim, S. B., Jeong, S. Y., Kim, W. H., & Park, J. G. (2005). Metabotropic glutamate receptor 4 expression in colorectal carcinoma and its prognostic significance. *Clin Cancer Res, 11*(9), 3288-3295. doi: 10.1158/1078-0432.CCR-04-1912

Constantinides, P. G., Jones, P. A., & Gevers, W. (1977). Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. *Nature, 267*(5609), 364-366.

Conticello, C., Adamo, L., Giuffrida, R., Vicari, L., Zeuner, A., Eramo, A., . . . De Maria, R. (2007). Proteasome inhibitors synergize with tumor necrosis factor-related apoptosis-induced ligand to induce anaplastic thyroid carcinoma cell death. *J Clin Endocrinol Metab, 92*(5), 1938-1942. doi: 10.1210/jc.2006-2157

Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C. L., & Feinberg, A. P. (2002). Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res, 62*(22), 6442-6446.

Dai, Y. N., Wang, J. H., Zhu, J. Z., Lin, J. Q., Yu, C. H., & Li, Y. M. (2015). Angiotensin-converting enzyme inhibitors/angiotensin receptor blockers therapy and colorectal cancer: a systematic review and meta-analysis. *Cancer Causes Control, 26*(9), 1245-1255. doi: 10.1007/s10552-015-0617-1

Darnell, J. E., Jr. (2002). Transcription factors as targets for cancer therapy. *Nat Rev Cancer, 2*(10), 740-749. doi: 10.1038/nrc906

De, S., Shaknovich, R., Riester, M., Elemento, O., Geng, H., Kormaksson, M., . . . Michor, F. (2013). Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet, 9*(1), e1003137. doi: 10.1371/journal.pgen.1003137

De Spiegeleer, B., Verbeke, F., D'Hondt, M., Hendrix, A., Van De Wiele, C., Burvenich, C., . . . Wynendaele, E. (2015). The quorum sensing peptides PhrG, CSP and EDF promote angiogenesis and invasion of breast cancer cells in vitro. *PLoS One, 10*(3), e0119471. doi: 10.1371/journal.pone.0119471

Denlinger, C. S., & Barsevick, A. M. (2009). The challenges of colorectal cancer survivorship. *J Natl Compr Canc Netw, 7*(8), 883-893; quiz 894.

Dorsam, R. T., & Gutkind, J. S. (2007). G-protein-coupled receptors and cancer. *Nat Rev Cancer, 7*(2), 79-94. doi: 10.1038/nrc2069

Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics, 11*, 587. doi: 10.1186/1471-2105-11-587

Dvorak, H. F. (2002). Vascular permeability factor/vascular endothelial growth factor: a critical cytokine in tumor angiogenesis and a potential target for diagnosis and therapy. *J Clin Oncol, 20*(21), 4368-4380.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene, 21*(35), 5400-5413. doi: 10.1038/sj.onc.1205651

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet, 8*(4), 286-298. doi: 10.1038/nrg2005

Esteller, M. (2008). Epigenetics in cancer. *N Engl J Med, 358*(11), 1148-1159. doi: 10.1056/NEJMra072067

Fan, S., & Zhang, X. (2009). CpG island methylation pattern in different human tissues and its correlation with gene expression. *Biochem Biophys Res Commun, 383*(4), 421-425. doi: 10.1016/j.bbrc.2009.04.023

Fang, J. Y., & Richardson, B. C. (2005). The MAPK signalling pathways and colorectal cancer. *Lancet Oncol, 6*(5), 322-327. doi: 10.1016/S1470-2045(05)70168-6

Fang, W. T., Fan, C. C., Li, S. M., Jang, T. H., Lin, H. P., Shih, N. Y., . . . Jiang, S. S. (2014). Downregulation of a putative tumor suppressor BMP4 by SOX2 promotes growth of lung squamous cell carcinoma. *Int J Cancer, 135*(4), 809-819. doi: 10.1002/ijc.28734

Feigin, M. E. (2013). Harnessing the genome for characterization of G-protein coupled receptors in cancer pathogenesis. *FEBS J, 280*(19), 4729-4738. doi: 10.1111/febs.12473

Feinberg, A. P., Ohlsson, R., & Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nat Rev Genet, 7*(1), 21-33. doi: nrg1748 [pii]

10.1038/nrg1748

Feinberg, A. P., & Tycko, B. (2004). The history of cancer epigenetics. *Nat Rev Cancer, 4*(2), 143-153. doi: 10.1038/nrc1279

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., . . . Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer, 136*(5), E359-386. doi: 10.1002/ijc.29210

Fleischer, T., Edvardsen, H., Solvang, H. K., Daviaud, C., Naume, B., Borresen-Dale, A. L., . . . Tost, J. (2014). Integrated analysis of high-resolution DNA methylation profiles, gene expression, germline genotypes and clinical end points in breast cancer patients. *Int J Cancer, 134*(11), 2615-2625. doi: 10.1002/ijc.28606

Fraga, M. F., Herranz, M., Espada, J., Ballestar, E., Paz, M. F., Ropero, S., . . . Esteller, M. (2004). A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors. *Cancer Res, 64*(16), 5527-5534. doi: 10.1158/0008-5472.CAN-03-4061

Franco, R., Schoneveld, O., Georgakilas, A. G., & Panayiotidis, M. I. (2008). Oxidative stress, DNA methylation and carcinogenesis. *Cancer Lett, 266*(1), 6-11. doi: 10.1016/j.canlet.2008.02.026

S0304-3835(08)00139-0 [pii]

Friedman, J. M., Liang, G., Liu, C. C., Wolff, E. M., Tsai, Y. C., Ye, W., . . . Jones, P. A. (2009). The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res, 69*(6), 2623-2629. doi: 10.1158/0008-5472.CAN-08-3114

Fulton, D. L., Sundararajan, S., Badis, G., Hughes, T. R., Wasserman, W. W., Roach, J. C., & Sladek, R. (2009). TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol, 10*(3), R29. doi: 10.1186/gb-2009-10-3-r29

Galm, O., Herman, J. G., & Baylin, S. B. (2006). The fundamental role of epigenetics in hematopoietic malignancies. *Blood Rev, 20*(1), 1-13. doi: 10.1016/j.blre.2005.01.006

Gehlenborg, N. (2016). UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets R package version 1.1.0. from https://CRAN.R-project.org/package=UpSetR

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., . . . Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74. doi: 10.1038/nature15393

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., . . . Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol, 5*(10), R80. doi: 10.1186/gb-2004-5-10-r80

Gervin, K., Vigeland, M. D., Mattingsdal, M., Hammero, M., Nygard, H., Olsen, A. O., . . . Lyle, R. (2012). DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet, 8*(1), e1002454. doi: 10.1371/journal.pgen.1002454

Gevaert, O., Tibshirani, R., & Plevritis, S. K. (2015). Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biol, 16*, 17. doi: 10.1186/s13059-014-0579-8

Gilabert, M., & Prebet, T. (2012). Acute leukemia arising after radioiodine treatment for thyroid cancer. *Haematologica, 97*(8), e28-29; author reply e30. doi: 10.3324/haematol.2012.067454

Giovannucci, E. (2001). Insulin, insulin-like growth factors and colon cancer: a review of the evidence. *J Nutr, 131*(11 Suppl), 3109S-3120S.

Golubovskaya, V. M., Kweh, F. A., & Cance, W. G. (2009). Focal adhesion kinase and cancer. *Histol Histopathol, 24*(4), 503-510.

Group, U. S. C. S. W. (2014). United States Cancer Statistics: 1999-2011 Incidence and Mortality Web-based Report, from www.cdc.gov/uscs

Gunes, E. G., Pinarbasi, E., Pinarbasi, H., & Silig, Y. (2009). Strong association between lung cancer and the AXIN2 polymorphism. *Mol Med Rep, 2*(6), 1029-1035. doi: 10.3892/mmr_00000210

Guo, S., Yan, F., Xu, J., Bao, Y., Zhu, J., Wang, X., . . . Wang, J. (2015). Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC). *Clin Epigenetics, 7*(1), 3. doi: 10.1186/s13148-014-0035-3

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70.

Heist, R. S., Sequist, L. V., & Engelman, J. A. (2012). Genetic changes in squamous cell lung cancer: a review. *J Thorac Oncol, 7*(5), 924-933. doi: 10.1097/JTO.0b013e31824cc334

Hejmadi, M. (2010). *Introduction to cancer biology*: Bookboon.

Herath, N. I., & Boyd, A. W. (2010). The role of Eph receptors and ephrin ligands in colorectal cancer. *Int J Cancer, 126*(9), 2003-2011. doi: 10.1002/ijc.25147

Herman, J. G., & Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med, 349*(21), 2042-2054. doi: 10.1056/NEJMra023075

Herman, J. G., Latif, F., Weng, Y., Lerman, M. I., Zbar, B., Liu, S., . . . et al. (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A, 91*(21), 9700-9704.

Hovestadt, V., Jones, D. T., Picelli, S., Wang, W., Kool, M., Northcott, P. A., . . . Lichter, P. (2014). Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature, 510*(7506), 537-541. doi: 10.1038/nature13268

Ideker, T., Ozier, O., Schwikowski, B., & Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics, 18 Suppl 1*, S233-240.

Irby, R. B., & Yeatman, T. J. (2000). Role of Src expression and activation in human cancer. *Oncogene, 19*(49), 5636-5642. doi: 10.1038/sj.onc.1203912

Jhala, N. C., Vickers, S. M., Argani, P., & McDonald, J. M. (2005). Regulators of apoptosis in cholangiocarcinoma. *Arch Pathol Lab Med, 129*(4), 481-486. doi: 10.1043/1543-2165(2005)129<481:ROAIC>2.0.CO;2

Jia, P., Liu, Y., & Zhao, Z. (2012). Integrative pathway analysis of genome-wide association studies and gene expression data in prostate cancer. *BMC Syst Biol, 6 Suppl 3*, S13. doi: 10.1186/1752-0509-6-S3-S13

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics, 8*(1), 118-127. doi: 10.1093/biostatistics/kxj037

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet, 3*(6), 415-428. doi: 10.1038/nrg816

nrg816 [pii]

Jones, P. A., & Baylin, S. B. (2007). The epigenomics of cancer. *Cell, 128*(4), 683-692. doi: S0092-8674(07)00127-4 [pii]

10.1016/j.cell.2007.01.029

Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nat Genet, 21*(2), 163-167. doi: 10.1038/5947

Jones, P. A., & Martienssen, R. (2005). A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. *Cancer Res, 65*(24), 11241-11246. doi: 65/24/11241 [pii]

10.1158/0008-5472.CAN-05-3865

Joshi, S. N., & Gardner, J. D. (1996). Gastrin and colon cancer: a unifying hypothesis. *Dig Dis, 14*(6), 334-344.

Jung-Hynes, B., Huang, W., Reiter, R. J., & Ahmad, N. (2010). Melatonin resynchronizes dysregulated circadian rhythm circuitry in human prostate cancer cells. *J Pineal Res, 49*(1), 60-68. doi: 10.1111/j.1600-079X.2010.00767.x

Kamburov, A., Stelzl, U., Lehrach, H., & Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res, 41*(Database issue), D793-800. doi: 10.1093/nar/gks1055

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. doi: 10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res, 44*(D1), D457-462. doi: 10.1093/nar/gkv1070

Kanemoto, M., Shirahata, M., Nakauma, A., Nakanishi, K., Taniguchi, K., Kukita, Y., . . . Kato, K. (2014). Prognostic prediction of glioblastoma by quantitative assessment of the methylation status of the entire MGMT promoter region. *BMC Cancer, 14*, 641. doi: 10.1186/1471-2407-14-641

Khor, T. O., Fuentes, F., Shu, L., Paredes-Gonzalez, X., Yang, A. Y., Liu, Y., . . . Kong, A. N. (2014). Epigenetic DNA methylation of antioxidative stress regulator NRF2 in human prostate cancer. *Cancer Prev Res (Phila), 7*(12), 1186-1197. doi: 10.1158/1940-6207.CAPR-14-0127

1940-6207.CAPR-14-0127 [pii]

Kim, J. H., Karnovsky, A., Mahavisno, V., Weymouth, T., Pande, M., Dolinoy, D. C., . . . Sartor, M. A. (2012). LRpath analysis reveals common pathways dysregulated via DNA methylation across cancer types. *BMC Genomics, 13*, 526. doi: 10.1186/1471-2164-13-526

Kimura, E. T., Nikiforova, M. N., Zhu, Z., Knauf, J. A., Nikiforov, Y. E., & Fagin, J. A. (2003). High prevalence of BRAF mutations in thyroid cancer: genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res, 63*(7), 1454-1457.

Klein, I., & Ojamaa, K. (2001). Thyroid hormone and the cardiovascular system. *N Engl J Med, 344*(7), 501-509. doi: 10.1056/NEJM200102153440707

Kou, Y., Zhang, S., Chen, X., & Hu, S. (2015). Gene expression profile analysis of colorectal cancer to investigate potential mechanisms using bioinformatics. *Onco Targets Ther, 8*, 745-752. doi: 10.2147/OTT.S78974

Kouvaraki, M. A., Shapiro, S. E., Perrier, N. D., Cote, G. J., Gagel, R. F., Hoff, A. O., . . . Evans, D. B. (2005). RET proto-oncogene: a review and update of genotype-phenotype correlations in hereditary medullary thyroid cancer and associated endocrine tumors. *Thyroid, 15*(6), 531-544. doi: 10.1089/thy.2005.15.531

Kramer, H. B., Lai, C. F., Patel, H., Periyasamy, M., Lin, M. L., Feller, S. M., . . . Buluwela, L. (2016). LRH-1 drives colon cancer cell growth by repressing the expression of the CDKN1A gene in a p53-dependent manner. *Nucleic Acids Res, 44*(2), 582-594. doi: 10.1093/nar/gkv948

Krueger, K. E., & Srivastava, S. (2006). Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics. *Mol Cell Proteomics, 5*(10), 1799-1810. doi: 10.1074/mcp.R600009-MCP200

Lanzetti, L., & Di Fiore, P. P. (2008). Endocytosis and cancer: an 'insider' network with dangerous liaisons. *Traffic, 9*(12), 2011-2021. doi: 10.1111/j.1600-0854.2008.00816.x

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., . . . Wei, C. L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res, 20*(3), 320-331. doi: 10.1101/gr.101907.109

Lee, S. T., & Wiemels, J. L. (2016). Genome-wide CpG island methylation and intergenic demethylation propensities vary among different tumor sites. *Nucleic Acids Res, 44*(3), 1105-1117. doi: 10.1093/nar/gkv1038

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi: 10.1186/1471-2105-12-323

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., . . . Parsons, R. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science, 275*(5308), 1943-1947.

Li, M., Balch, C., Montgomery, J. S., Jeong, M., Chung, J. H., Yan, P., . . . Nephew, K. P. (2009). Integrated analysis of DNA methylation and gene expression reveals specific signaling pathways associated with platinum resistance in ovarian cancer. *BMC Med Genomics, 2*, 34. doi: 10.1186/1755-8794-2-34

Linehan, W. M., Srinivasan, R., & Schmidt, L. S. (2010). The genetic basis of kidney cancer: a metabolic disease. *Nat Rev Urol, 7*(5), 277-285. doi: 10.1038/nrurol.2010.47

Lu, P., Weaver, V. M., & Werb, Z. (2012). The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol, 196*(4), 395-406. doi: 10.1083/jcb.201102147

Lucas, S. D., Karlsson-Parra, A., Nilsson, B., Grimelius, L., Akerstrom, G., Rastad, J., & Juhlin, C. (1996). Tumor-specific deposition of immunoglobulin G and complement in papillary thyroid carcinoma. *Hum Pathol, 27*(12), 1329-1335.

Ma, X., Gao, Y., Fan, Y., Ni, D., Zhang, Y., Chen, W., . . . Zhang, X. (2013). Overexpression of E2F1 promotes tumor malignancy and correlates with TNM stages in clear cell renal cell carcinoma. *PLoS One, 8*(9), e73436. doi: 10.1371/journal.pone.0073436

Major, J. M., Pollak, M. N., Snyder, K., Virtamo, J., & Albanes, D. (2010). Insulin-like growth factors and risk of kidney cancer in men. *Br J Cancer, 103*(1), 132-135. doi: 10.1038/sj.bjc.6605722

Mansuroglu, T., Baumhoer, D., Dudas, J., Haller, F., Cameron, S., Lorf, T., . . . Ramadori, G. (2009). Expression of stem cell factor receptor c-kit in human nontumoral and tumoral hepatic cells. *Eur J Gastroenterol Hepatol, 21*(10), 1206-1211. doi: 10.1097/MEG.0b013e328317f4ef

Marcucci, G., Yan, P., Maharry, K., Frankhouser, D., Nicolet, D., Metzeler, K. H., . . . Bloomfield, C. D. (2014). Epigenetics meets genetics in acute myeloid leukemia: clinical impact of a novel seven-gene score. *J Clin Oncol, 32*(6), 548-556. doi: 10.1200/JCO.2013.50.6337

Marie Laure Delignette-Muller, C. D. (2015). An R package for fitting distributions. *Journal of Statistical Software, 64*(4), 1-34.

McGregor, L. M., McCune, B. K., Graff, J. R., McDowell, P. R., Romans, K. E., Yancopoulos, G. D., . . . Nelkin, B. D. (1999). Roles of trk family neurotrophin receptors in medullary thyroid carcinoma development and progression. *Proc Natl Acad Sci U S A, 96*(8), 4540-4545.

McGuire, S. (2016). World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Adv Nutr, 7*(2), 418-419. doi: 10.3945/an.116.012211

7/2/418 [pii]

Mohammadi-asl, J., Larijani, B., Khorgami, Z., Tavangar, S. M., Haghpanah, V., Kheirollahi, M., & Mehdipour, P. (2011). Qualitative and quantitative promoter hypermethylation patterns of the P16, TSHR, RASSF1A and RARbeta2 genes in papillary thyroid carcinoma. *Med Oncol, 28*(4), 1123-1128. doi: 10.1007/s12032-010-9587-z

Molloy, N. H., Read, D. E., & Gorman, A. M. (2011). Nerve growth factor in cancer cell death and survival. *Cancers (Basel), 3*(1), 510-530. doi: 10.3390/cancers3010510

Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., & Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics, 30*(3), 428-430. doi: 10.1093/bioinformatics/btt684

btt684 [pii]

Najdi, R., Holcombe, R. F., & Waterman, M. L. (2011). Wnt signaling and colon carcinogenesis: beyond APC. *J Carcinog, 10*, 5. doi: 10.4103/1477-3163.78111

Niculescu, F., Rus, H. G., Retegan, M., & Vlaicu, R. (1992). Persistent complement activation on tumor cells in breast cancer. *Am J Pathol, 140*(5), 1039-1043.

Okamoto, K., Tajima, H., Ohta, T., Nakanuma, S., Hayashi, H., Nakagawara, H., . . . Iseki, S. (2010). Angiotensin II induces tumor progression and fibrosis in intrahepatic cholangiocarcinoma through an interaction with hepatic stellate cells. *Int J Oncol, 37*(5), 1251-1259.

Oliveros, J. C. (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams., from http://bioinfogp.cnb.csic.es/tools/venny/index.html

Owens, L. V., Xu, L., Dent, G. A., Yang, X., Sturge, G. C., Craven, R. J., & Cance, W. G. (1996). Focal adhesion kinase as a marker of invasive potential in differentiated human thyroid cancer. *Ann Surg Oncol, 3*(1), 100-105.

Ozcan, A., Shen, S. S., Hamilton, C., Anjana, K., Coffey, D., Krishnan, B., & Truong, L. D. (2011). PAX 8 expression in non-neoplastic tissues, primary tumors, and metastatic tumors: a comprehensive immunohistochemical study. *Mod Pathol, 24*(6), 751-764. doi: 10.1038/modpathol.2011.3

Ozer, B., & Sezerman, O. (2015). A novel analysis strategy for integrating methylation and expression data reveals core pathways for thyroid cancer aetiology. *BMC Genomics, 16 Suppl 12*, S7. doi: 10.1186/1471-2164-16-S12-S7

Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., . . . Meyerson, M. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science, 304*(5676), 1497-1500. doi: 10.1126/science.1099314

Patsos, H. A., Greenhough, A., Hicks, D. J., Al Kharusi, M., Collard, T. J., Lane, J. D., . . . Williams, A. C. (2010). The endogenous cannabinoid, anandamide, induces COX-2-dependent cell death in apoptosis-resistant colon cancer cells. *Int J Oncol, 37*(1), 187-193.

Paziewska, A., Dabrowska, M., Goryca, K., Antoniewicz, A., Dobruch, J., Mikula, M., . . . Ostrowski, J. (2014). DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy. *Br J Cancer, 111*(4), 781-789. doi: 10.1038/bjc.2014.337

Prickett, T. D., & Samuels, Y. (2012). Molecular pathways: dysregulated glutamatergic signaling pathways in cancer. *Clin Cancer Res, 18*(16), 4240-4246. doi: 10.1158/1078-0432.CCR-11-1217

Priolo, C., Pyne, S., Rose, J., Regan, E. R., Zadra, G., Photopoulos, C., . . . Loda, M. (2014). AKT1 and MYC induce distinctive metabolic fingerprints in human prostate cancer. *Cancer Res, 74*(24), 7198-7204. doi: 10.1158/0008-5472.CAN-14-1490

QIAGEN. (2016). QIAGEN - Oncogenes and Tumor Suppressor Genes RT2 Profiler PCR arrays Retrieved 16/02/2016, 2016, from https://www.qiagen.com/gb/shop/pcr/primer-sets/rt2-profiler-pcr-arrays?catno=PAHS-502Z#resources

R Development Core Team. (2014). R: A language and environment for statistical computing. Vienna,Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi: 10.1093/bioinformatics/btp616

Roy, I., Zimmerman, N. P., Mackinnon, A. C., Tsai, S., Evans, D. B., & Dwinell, M. B. (2014). CXCL12 chemokine expression suppresses human pancreatic cancer growth and metastasis. *PLoS One, 9*(3), e90400. doi: 10.1371/journal.pone.0090400

Rzeski, W., Ikonomidou, C., & Turski, L. (2002). Glutamate antagonists limit tumor growth. *Biochem Pharmacol, 64*(8), 1195-1200.

Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., . . . Lancet, D. (2010). GeneCards Version 3: the human gene integrator. *Database (Oxford), 2010*, baq020. doi: 10.1093/database/baq020

Saharinen, P., Tammela, T., Karkkainen, M. J., & Alitalo, K. (2004). Lymphatic vasculature: development, molecular regulation and role in tumor metastasis and inflammation. *Trends Immunol, 25*(7), 387-395. doi: 10.1016/j.it.2004.05.003

Saito, Y., Liang, G., Egger, G., Friedman, J. M., Chuang, J. C., Coetzee, G. A., & Jones, P. A. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell, 9*(6), 435-443. doi: 10.1016/j.ccr.2006.04.020

Sanz-Pamplona, R., Berenguer, A., Cordero, D., Mollevi, D. G., Crous-Bou, M., Sole, X., . . . Moreno, V. (2014). Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. *Mol Cancer, 13*, 46. doi: 10.1186/1476-4598-13-46

Schroder, M. S., Culhane, A. C., Quackenbush, J., & Haibe-Kains, B. (2011). survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics, 27*(22), 3206-3208. doi: 10.1093/bioinformatics/btr511

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504. doi: 10.1101/gr.1239303

Sharma, S., Kelly, T. K., & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis, 31*(1), 27-36. doi: 10.1093/carcin/bgp220

Sherr, C. J. (2000). Cell cycle control and cancer. *Harvey Lect, 96*, 73-92.

Shulman, G. I., Ladenson, P. W., Wolfe, M. H., Ridgway, E. C., & Wolfe, R. R. (1985). Substrate cycling between gluconeogenesis and glycolysis in euthyroid, hypothyroid, and hyperthyroid man. *J Clin Invest, 76*(2), 757-764. doi: 10.1172/JCI112032

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., . . . Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature, 511*(7511), 611-615. doi: 10.1038/nature13581

Soares, M. J., Pinto, M., Henrique, R., Vieira, J., Cerveira, N., Peixoto, A., . . . Teixeira, M. R. (2009). CSF1R copy number changes, point mutations, and RNA and protein overexpression in renal cell carcinomas. *Mod Pathol, 22*(6), 744-752. doi: 10.1038/modpathol.2009.43

Spano, J. P., Fagard, R., Soria, J. C., Rixe, O., Khayat, D., & Milano, G. (2005). Epidermal growth factor receptor signaling in colorectal cancer: preclinical data and therapeutic perspectives. *Ann Oncol, 16*(2), 189-194. doi: 10.1093/annonc/mdi057

Stern, D. F. (2000). Tyrosine kinase signalling in breast cancer: ErbB family receptor tyrosine kinases. *Breast Cancer Res, 2*(3), 176-183.

Stewart, D. A., Cooper, C. R., & Sikes, R. A. (2004). Changes in extracellular matrix (ECM) and ECM-associated proteins in the metastatic progression of prostate cancer. *Reprod Biol Endocrinol, 2*, 2. doi: 10.1186/1477-7827-2-2

Stuver, S. O., Kuper, H., Tzonou, A., Lagiou, P., Spanos, E., Hsieh, C. C., . . . Trichopoulos, D. (2000). Insulin-like growth factor 1 in hepatocellular carcinoma and metastatic liver cancer in men. *Int J Cancer, 87*(1), 118-121.

Swinnen, J. V., Vanderhoydonc, F., Elgamal, A. A., Eelen, M., Vercaeren, I., Joniau, S., . . . Verhoeven, G. (2000). Selective activation of the fatty acid synthesis pathway in human prostate cancer. *Int J Cancer, 88*(2), 176-179.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452. doi: 10.1093/nar/gku1003

Szmida, E., Karpinski, P., Leszczynski, P., Sedziak, T., Kielan, W., Ostasiewicz, P., & Sasiadek, M. M. (2015). Aberrant methylation of ERBB pathway genes in sporadic colorectal cancer. *J Appl Genet, 56*(2), 185-192. doi: 10.1007/s13353-014-0253-6

Tada, M., Omata, M., & Ohto, M. (1992). High incidence of ras gene mutation in intrahepatic cholangiocarcinoma. *Cancer, 69*(5), 1115-1118.

Tang, Y., Choi, E. J., Cheong, S. H., Hwang, Y. J., Arokiyaraj, S., Park, P. J., . . . Kim, E. K. (2015). Effect of taurine on prostate-specific antigen level and migration in human prostate cancer cells. *Adv Exp Med Biol, 803*, 203-214. doi: 10.1007/978-3-319-15126-7_18

Taskesen, E., Havermans, M., van Lom, K., Sanders, M. A., van Norden, Y., Bindels, E., . . . Delwel, R. (2014). Two splice-factor mutant leukemia subgroups uncovered at the boundaries of MDS and AML using combined gene expression and DNA-methylation profiling. *Blood, 123*(21), 3327-3335. doi: 10.1182/blood-2013-07-512855

Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., . . . Gerald, W. L. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell, 18*(1), 11-22. doi: 10.1016/j.ccr.2010.05.026

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics, 29*(2), 189-196. doi: 10.1093/bioinformatics/bts680

Tian, X., Sun, D., Zhao, S., Xiong, H., & Fang, J. (2015). Screening of potential diagnostic markers and therapeutic targets against colorectal cancer. *Onco Targets Ther, 8*, 1691-1699. doi: 10.2147/OTT.S81621

Tsouko, E., Khan, A. S., White, M. A., Han, J. J., Shi, Y., Merchant, F. A., . . . Frigo, D. E. (2014). Regulation of the pentose phosphate pathway by an androgen receptor-mTOR-mediated mechanism and its role in prostate cancer cell growth. *Oncogenesis, 3*, e103. doi: 10.1038/oncsis.2014.18

van Eijk, K. R., de Jong, S., Boks, M. P., Langeveld, T., Colas, F., Veldink, J. H., . . . Ophoff, R. A. (2012). Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics, 13*, 636. doi: 10.1186/1471-2164-13-636

Vieira, J. M., Santos, S. C., Espadinha, C., Correia, I., Vag, T., Casalou, C., . . . Leite, V. (2005). Expression of vascular endothelial growth factor (VEGF) and its receptors in thyroid carcinomas of follicular origin: a potential autocrine loop. *Eur J Endocrinol, 153*(5), 701-709. doi: 10.1530/eje.1.02009

Vivanco, I., & Sawyers, C. L. (2002). The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev Cancer, 2*(7), 489-501. doi: 10.1038/nrc839

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science, 339*(6127), 1546-1558. doi: 10.1126/science.1235122

Wan, J., Oliver, V. F., Wang, G., Zhu, H., Zack, D. J., Merbs, S. L., & Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics, 16*, 49. doi: 10.1186/s12864-015-1271-4

Wang, Y., Li, G. W., & Wen, B. G. (2005). [Loss of IGF2 imprinting in colorectal cancer]. *Sheng Li Ke Xue Jin Zhan, 36*(1), 71-73.

Wei, J. H., Haddad, A., Wu, K. J., Zhao, H. W., Kapur, P., Zhang, Z. L., . . . Luo, J. H. (2015). A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma. *Nat Commun, 6*, 8699. doi: 10.1038/ncomms9699

Wongpaiboonwattana, W., Tosukhowong, P., Dissayabutra, T., Mutirangura, A., & Boonla, C. (2013). Oxidative stress induces hypomethylation of LINE-1 and hypermethylation of the RUNX3 promoter in a bladder cancer cell line. *Asian Pac J Cancer Prev, 14*(6), 3773-3778.

Wu, H., Chen, Y., Liang, J., Shi, B., Wu, G., Zhang, Y., . . . Shang, Y. (2005). Hypomethylation-linked activation of PAX2 mediates tamoxifen-stimulated endometrial carcinogenesis. *Nature, 438*(7070), 981-987. doi: 10.1038/nature04225

Wu, Y., Wang, X., Wu, F., Huang, R., Xue, F., Liang, G., . . . Huang, Y. (2012). Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing. *PLoS One, 7*(8), e41001. doi: 10.1371/journal.pone.0041001

Xu, L. H., Yang, X., Bradham, C. A., Brenner, D. A., Baldwin, A. S., Jr., Craven, R. J., & Cance, W. G. (2000). The focal adhesion kinase suppresses transformation-associated, anchorage-independent apoptosis in human breast cancer cells. Involvement of death receptor-related signaling pathways. *J Biol Chem, 275*(39), 30597-30604. doi: 10.1074/jbc.M910027199

Yang, A. S., Estecio, M. R., Garcia-Manero, G., Kantarjian, H. M., & Issa, J. P. (2003). Comment on "Chromosomal instability and tumors promoted by DNA hypomethylation" and "Induction of tumors in nice by genomic hypomethylation". *Science, 302*(5648), 1153; author reply 1153. doi: 10.1126/science.1089523

Yang, J., & Weinberg, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell, 14*(6), 818-829. doi: 10.1016/j.devcel.2008.05.009

Yilmaz, H. H., Yazihan, N., Tunca, D., Sevinc, A., Olcayto, E. O., Ozgul, N., & Tuncer, M. (2011). Cancer trends and incidence and mortality patterns in Turkey. *Jpn J Clin Oncol, 41*(1), 10-16. doi: 10.1093/jjco/hyq075

hyq075 [pii]

Yoo, C. B., & Jones, P. A. (2006). Epigenetic therapy of cancer: past, present and future. *Nat Rev Drug Discov, 5*(1), 37-50. doi: 10.1038/nrd1930

Yoo, S., Takikawa, S., Geraghty, P., Argmann, C., Campbell, J., Lin, L., . . . Zhu, J. (2015). Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet, 11*(1), e1004898. doi: 10.1371/journal.pgen.1004898

Yoshiji, H., Noguchi, R., Ikenaka, Y., Kaji, K., Aihara, Y., & Fukui, H. (2011). Impact of renin-angiotensin system in hepatocellular carcinoma. *Curr Cancer Drug Targets, 11*(4), 431-441.

Zaykin, D. V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol, 24*(8), 1836-1841. doi: 10.1111/j.1420-9101.2011.02297.x

Zhao, M., Kim, P., Mitra, R., Zhao, J., & Zhao, Z. (2016). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res, 44*(D1), D1023-1031. doi: 10.1093/nar/gkv1268

Zhu, X., Zhu, Y. J., Kim, D. W., Meltzer, P., & Cheng, S. Y. (2014). Activation of integrin-ERBB2 signaling in undifferentiated thyroid cancer. *Am J Cancer Res, 4*(6), 776-788.

8. APPENDICES

Appendix 1

45 genes that are shared by at least 3 types of cancers with more than 25% methylation change. List of significantly methylated genes (FDR<0.05) for thyroid, breast, colon and prostate cancers. Only the genes that are shared by more than two cancer types are shown at this table. In addition, descriptions associated with each gene are also added to table using Genecards suite.

GeneName	Brca_Meth	Coad_Meth	Prad_Meth	Thca_Meth	Gene Description
LAYN	0.251	0.363	0.274	-	Receptor for hyaluronate
RARRES2	0.389	0.278	0.289	-	Regulates adipocyte differentiation. angiogenesis
MYLK	0.278	0.394	0.284	-	Immunoglobulin gene superfamily
GPR115	-0.310	-0.356	0.268	-0.326	
MS4A15	-0.359	-0.304	-	-0.253	May be involved in signal transduction
USP44	0.353	0.546	0.342	-	Regulatory role in the spindle assembly checkpoint or mitotic checkpoint
CPEB1	0.351	0.450	0.275	-	Member of cytoplasmic polyadenylation element binding protein family
ZNF454	0.349	0.357	0.283	-	May be involved in transcriptional regulation
KIAA1614	0.325	0.335	0.356	-	
ZNF662	0.342	0.253	0.261	-	May be involved in transcriptional regulation
GALR1	0.388	0.441	0.271	-	Receptor for the hormone galanin
GNAL	0.257	0.356	0.254	-	Modulators or transducers in various transmembrane signaling systems
CLIP4	0.346	0.409	0.343	-	CAP-GLY domain containing linker protein family
DPP6	0.269	0.463	0.268	-	Involved in the physiological processes of brain function
TRIM9	0.297	0.391	0.252	-	Role in regulation of neuronal functions and synaptic vesicle exocytosis
SFRP5	0.370	0.371	0.337	-	Regulating cell growth and differentiation
EFEMP1	0.378	0.322	0.293	-	Role in cell adhesion and migration
VWC2	0.364	0.454	0.281	-	Play a role in neural development. Promotes cell adhesion
LOC151174	0.266	0.263	0.256	-	
PDLIM4	0.263	0.325	0.263	-	Protein which may be involved in bone development
SCGB3A1	0.283	0.322	0.296	-	Potential growth inhibitory cytokine
SYNPO	0.429	-	0.261	-0.360	Essential for synaptic plasticity
FZD7	0.311	0.299	0.290	-	Receptor for Wnt proteins
FAT4	0.352	0.302	0.250	-	Regulation of planar cell polarity
SORBS2	0.320	-	0.349	0.362	Assembling of signaling complexes
CFL2	0.336	0.278	0.267	-	Controls actin polymerization in a pH-sensitive manner
DCHS2	0.297	0.308	0.255	-	Calcium-dependent cell-adhesion protein
NRG1	0.297	0.428	0.289	-	Signaling protein. involved in growth and development of organ systems
MEIS2	0.299	0.433	0.292	-	Involved in transcriptional regulation
FLNC	0.257	0.320	0.251	-	Central role in muscle cells
SLC7A14	0.432	0.270	0.260	-	Member of solute carrier family
COL8A1	-0.401	-0.482	-	-0.298	Collagen
EFEMP2	0.271	0.276	0.255	-	Elastic fiber formation and connective tissue development
C16orf5	0.409	0.264	0.354	-	Important p53/TP53-apoptotic effector

APBB1	0.343	0.463	0.270	-	Transcription coregulator; both coactivator and corepressor functions
TMEM51	-0.326	-	0.301	-0.277	Transmembrane protein
EVC2	0.311	0.456	0.331	-	Positive regulator of the hedgehog signaling pathway
MPPED2	0.280	0.269	0.347	-	Metallophosphoesterase; development of the nervous system
CNTN1	0.311	0.380	0.281	-	Formation of axon connections; immunoglobulin superfamily
SEMA6C	0.365	0.382	0.276	-	Semaphorin family; important role in neural regeneration
GPRC5B	0.419	0.428	0.410	-	Member of type 3 G protein-coupled receptor family
KCNB2	0.257	0.352	0.254	-	Mediates voltage-dependent potassium ion permeability of membranes
CLIP3	0.255	0.283	0.262	-	Cytoplasmic linker protein
HIF3A	0.295	0.252	0.254	-	Regulate many adaptive responses to low oxygen tension (hypoxia)
LOC100126784	-	0.415	0.279	-0.451	

Appendix 2

Table showing number of samples with decreasing and increasing copy numbers of driver genes. Only the genes which were detected as being affected by high methylation change in maximum proximity of 3 are included at the analysis.

	Gene Name	Number of Samples with Increased CN	Number of Samples with Decreased CN	Gene Information
THCA - 46 tumour samples	AGTR1	0	1	suppressor
	FGFR3	1	6	oncogene
	PPARG	0	4	suppressor
	RET	0	4	oncogene
CHOL - 9 tumour samples	FOXO1	0	0	suppressor
	PPARG	0	0	suppressor
	CXCL12	0	0	suppressor
	HRAS	0	2	oncogene
	KIT	0	1	oncogene
	AGTR1	0	0	suppressor
	CDK4	0	0	oncogene
	ZBTB16	0	0	suppressor
	FAS	0	0	suppressor
	NKX3-1	0	0	suppressor
	EPAS1	0	1	suppressor
	FGFR3	0	2	oncogene
	IGF1	0	0	suppressor
	NCOA4	0	1	suppressor
	PIK3R1	0	1	suppressor
	SRC	0	0	oncogene
COAD- 15 tumour samples	PRKCB	0	0	suppressor
	CXCL12	0	1	suppressor
	IGF1	0	0	suppressor
	AGTR1	0	0	suppressor
	MAPK10	0	0	suppressor
	CDKN1A	0	1	suppressor
	EDNRB	0	0	suppressor
	EPAS1	0	2	suppressor
	FOXD3	1	2	suppressor
KIRP - 23 tumour samples	FOXD3	0	0	suppressor
	FOXO1	0	0	suppressor
	MET	0	2	oncogene
	SRC	0	3	oncogene
	WT1	0	1	suppressor
	CSF1R	0	0	oncogene
	IGF1	0	0	suppressor
	E2F1	0	3	suppressor

LUSC - 7 tumuor samples	BMP4	1	2	suppressor
	DCC	0	0	suppressor
	EDNRB	0	1	suppressor
	EPAS1	0	3	suppressor
	FGFR3	1	0	oncogene
	PIK3R1	1	2	suppressor
	SMO	1	0	oncogene
	EGFR	3	1	oncogene
	AGTR1	5	0	suppressor
	PRKCB	0	0	suppressor
	CXCL12	0	1	suppressor
	IGF1	1	1	suppressor
	AXIN2	0	0	suppressor
	SPI1	4	0	suppressor
	CDK4	1	1	oncogene
Liver - 39 tumour samples	FOXD3	4	2	suppressor
	IGF1	0	2	suppressor
	SRC	1	3	oncogene
	CXCL12	2	3	suppressor
	NKX3-1	1	3	suppressor
	AGTR1	1	1	suppressor