A BIOINFORMATICS META-ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES IN COLORECTAL CANCER

by

SIMON KIT CHAN

B.Sc. First Class Honours, Cell and Molecular Biology,

Simon Fraser University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Bioinformatics)

THE UNIVERSITY OF BRITISH COLUMBIA

December 2007

Abstract

BACKGROUND: Elucidation of candidate colorectal cancer biomarkers often begins by comparing the expression profiles of cancerous and normal tissue by performing high throughput gene expression profiling. While many such studies have been performed, the resulting lists of differentially expressed genes tend to be inconsistent with each other, suggesting that there are some false positives and negatives. One logical solution to this problem is to determine the intersection of the lists of differentially expressed genes from independent studies. It is expected that genes that are biologically relevant to cancer tumorigenesis will be reported most often, while sporadically reported genes are due to the inherent biases and limitations of each of the profiling platforms used. However, the statistical significance of the observed intersection among many independent studies is usually not considered. **PURPOSE:** To address these issues, we developed a computational meta-analysis method that ranked differentially expressed genes based on the following criteria, which are presented in order of importance: the amount of intersection among studies, total tissue sample sizes, and average fold change in expression. We applied this meta-analysis method to 25 independent colorectal cancer profiling studies that compared cancer versus normal, adenoma versus normal, and cancer versus adenoma tissues. **RESULTS:** We observed that some genes were consistently reported as differentially expressed with a statistically significant frequency ($P < .0001$) in the cancer versus normal and adenoma versus normal comparisons,

but not in the cancer versus adenoma comparison. We performed a review of some of the high ranking candidates and determined that some have previously been shown to have diagnostic and/or prognostic utility in colorectal cancer. More interestingly, the meta-analysis method also identified genes that had yet to be tested and validated as biomarkers. Thus, these candidates are currently being validated at the protein level on colorectal tissue microarrays.

CONCLUSION: Our meta-analysis method identified genes that were consistently reported as differentially expressed. Besides identifying new biomarker candidates, our meta-analysis method also provides another filter to remove false positive genes from further consideration. In conclusion, the genes presented here will aid in the identification of highly sensitive and specific biomarkers in colorectal cancer.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I wish to acknowledge my senior supervisor, Dr. Steven Jones, and my committee member, Dr. Isabella Tai. In a letter to Robert Hooke, Sir Isaac Newton wrote "If I have seen further, it is by standing on ye shoulders of Giants." I am grateful to Dr. Jones and Dr. Tai for their patience and guidance throughout this thesis research. If I have gained anything from my graduate training, it is because they have kindly let me stand on their shoulders.

I would also like to acknowledge Dr. Mark Wilkinson for being on my committee, as well as for asking thought provoking questions at my committee meetings.

I wish to thank members of the Genome Sciences Centre where this thesis research was conducted. They provided much support and contributed to a friendly working environment. Special thanks to fellow graduate student, Obi Griffith, for the valuable insight he provided during our countless discussions. I am grateful to the CIHR/MSFHR Strategic Training Program in Bioinformatics for Health Research for providing financial support. Furthermore, the UBC Bioinformatics Coordinator, Sharon Ruschkowski, provided support in applying for funding and registration for classes.

Finally, I wish to thank my parents for their unconditional love and support.

Dedication

To my parents

Co-Authorship Statement

Dr. Steven Jones, Dr. Isabella Tai, Obi Griffith, and Simon Chan all made important intellectual contributions to this thesis research. Obi Griffith developed an earlier version of the meta-analysis method that was used in this thesis research and contributed a portion of the bioinformatics code. Simon Chan performed the data collection, curation, performed the analyses, produced the figures and tables, and wrote this document. Dr. Isabella Tai supervised the data collection portion of this thesis research and helped interpret results. Dr. Steven Jones conceived of this project, helped interpret the results, and provided financial support. Finally, Rhonda Oshanek edited the document for content and grammar.

Chapter 1: Introduction[1]

1.1 Introduction to Colorectal Cancer

Colorectal cancer is defined as cancerous growths in the colon, rectum, or appendix and represents the third most frequent cancer in both men and women in North America [1]. In Canada, colorectal cancer is the second most common cause of cancer-related death. This year, an estimated 20,800 Canadians will be diagnosed with colorectal cancer and approximately 8,700 will die of it [2]. When diagnosed and treated early, the majority of patients show favourable five year survival rates. Among patients diagnosed when the cancer is localized to the colon or rectum, 90% of patients show good survival rates, while the survival rate is 68% among patients diagnosed when the cancer has spread to the surrounding tissues. Unfortunately, approximately 19% of patients are diagnosed when the cancer has metastasized to other organs, such as the liver, and the corresponding survival rate is only 10% [3]. Thus, regular screening remains the most promising strategy to reducing the mortality rates associated with colorectal cancer.

---

[1] A version of each chapter was compiled into a manuscript that has been accepted for publication in the peer reviewed journal *Cancer Epidemiology, Biomarkers & Prevention*:

Chan SK, Griffith OL, Tai IT, Jones SJM. **Meta-analysis of Colorectal Cancer Gene Expression Profiling Studies Identifies Consistently Reported Candidate Biomarkers.** (2007). In press. *Cancer Epidemiol Biomarkers Prev.*

1.2 Stages of Colorectal Cancer

The transformation of healthy colon epithelia to neoplasia has provided an excellent system to study the histological and molecular changes that are associated with the progression of colorectal cancer [4, 5]. The first step in the development of colorectal cancer is the emergence of aberrant crypt foci (ACF), which are small lesions in the intestinal epithelium [6]. ACF are composed of cells of normal morphology or dysplastic cells, of which the latter are more likely to develop into benign polyps, also known as adenomas, which protrude into the lumen. By the age of 70, approximately 50% of men and women from the western world will have developed an adenoma polyp [5]. Approximately one in ten of these adenomas will progress into the carcinoma stage.

As for many other solid tumors, colorectal cancer is staged mainly on the basis of both the extent of invasion at the primary site and of the presence of metastasis in local lymph nodes and distant organs. The two most common colorectal cancer staging systems are Dukes [7] and TNM [8]. Both systems distinguish between tumors confined to the surface of the colon wall (Dukes A or Stage I), those penetrating the muscle wall but without metastasis (Dukes B or Stage II), those with metastasis to surrounding lymph nodes (Dukes C or Stage III), and finally, those that have metastasized to distant organs, such as the liver (Dukes D or Stave IV).

1.3 The Underlying Molecular Changes during the Progression of Normal to Cancer Tissue: The Wnt Signaling Pathway and the Adenomatous Polyposis Coli (APC) Gene

Molecular changes in colonic epithelium underlie the histological changes seen in the progression of normal to cancer tissue. One molecular pathway that has been implicated in the development of colorectal cancer is that of the Wnt signaling pathway. This pathway controls cellular division, cell-cell adhesion, and other activities by inducing the expression of genes like *myc*, *ccnd1*, and *mmp-7* [9-12]. This induction is mediated through the binding of the extracellular Wnt signal to the Frizzled and Low density lipoprotein binding protein (LBP) transmembrane proteins. In the absence of the Wnt signal, cytoplasmic β-catenin is destabilized through its interaction with a destruction complex, made up of scaffolding proteins Adenomatous Polyposis Coli (APC), axin (AXIN2), and gylogen synthase kinase 3β (GSK3β). When bound to the destruction complex, cytoplasmic β-catenin is ubquinated, leading to its proteasomal degradation. When the Wnt signal is present and binds to Frizzled and LBP, the destruction complex is inactivated via Dishevelled (DSH), thus allowing stabilized β-catenin to enter the nucleus and interact with DNA-binding proteins of the T-cell-factor (TCF) family. Finally, this complex induces expression of target genes by binding to their respective regulatory regions.

1.3.1 APC is the Gatekeeper to Colorectal Cancer

The *apc* gene and the effects of its mutation have been an area of great interest within the colorectal cancer research community. Inactivation of both alleles of *apc* has been located in dysplastic aberrant crypt foci, which are the earliest neoplastic lesions and an early precursor to adenomas [5]. Furthermore, mutations in this tumor suppressor gene have been found in up to 80% of colorectal cancer samples [13]. Thus, *apc* has often been described as the "gatekeeper" to colorectal cancer, as its inactivation leads to the initiation of the disease [14].

As described, the APC protein is a part of the destruction complex within the Wnt pathway. Inactivation of both alleles of *apc* renders the destruction complex ineffective at promoting the degradation of cytoplasmic β-catenin, thus allowing it to accumulate and enter the cell nucleus, where it binds with TCF and together induce the expression of the target genes. Thus, the Wnt signal transduction pathway becomes constitutively active, with the net result being an increase in cell division of the colonic epithelial cells. Furthermore, recent studies have also shown that even one mutant allele of *apc* can contribute to the chromosome instability of colorectal tumor cells [15, 16]. APC protein has been implicated in the proper formation and function of mitotic spindles, as mutations in this gene have resulted in cells with chromosome segregation errors [15, 17, 18]. That only one mutant allele of *apc* can result in these errors suggests that the mutant

allele has a dominant negative effect on the remaining wildtype allele. Consistent with this hypothesis, a recent study demonstrated that a mutant APC protein forms a hetero-oligomer with that of the endogenous APC protein, thus interfering with its proper functioning [16].

1.3.2 K-ras Mutations in Intermediate Stage Adenomas

After development of adenomas from dysplastic ACF, mutations in other genes contribute to colorectal cancer progression. For example, approximately 50% of intermediate-staged adenomas have activating mutations in the *k-ras* oncogene [4, 19]. *K-ras* is part of a family of three genes, the other members being *h-ras* and *n-ras*, and encodes a small GTPase involved in the transduction of external stimuli across the plasma membrane to the effector molecules that promote cell division. An activation mutation in just one allele of this gene results in the inability of *k-ras* to be inactivated, thus resulting in uncontrolled cell division and malignant transformation [19]. Furthermore, mutated *k-ras* has been shown to promote local cell invasion by inducing the expression of metalloproteases, such as *mmp-7* [20], to breakdown the underlying basement membrane of the epithelial cells, thus releasing tumor cells from the primary tumor [21].

1.3.3 Loss of Heterozygosity on 18q is associated with Late Stage Adenomas

1.3.3.1 The Deleted in Colorectal Cancer (DCC) Gene is a Candidate Tumor Suppressor on 18q

Loss of heterozygosity (LOH) on 18q, found in approximately 70% of colorectal cancer samples, is associated with the progression of intermediate-stage adenomas to late-stage adenomas [4, 5, 22]. LOH has been implicated as an important mechanism for the inactivation of tumor suppressor genes [23]. Deleted in Colorectal Cancer (DCC) is one candidate tumor suppressor gene that is in the deleted region [24]. DCC functions as part of a receptor complex for the netrin 1 (NTN1) protein [25-27]. The netrins are a family of secreted proteins with roles in directing the growth and orientation of axons in the developing nervous system [28, 29]. While an obvious role for DCC and NTN1 in the intestine is not evident [28, 30], some studies have suggested that they have roles in regulating apoptosis, in which the expression of DCC in the absence of NTN1 induces apoptosis, while in the presence of NTN1, this expression is blocked [31-33]. Thus, DCC has been proposed to be a member of the Dependence Receptor Family. These receptors promote a certain cellular state in the presence of the ligand, but in its absence, the receptors promote apoptosis [34], which suggests a possible role for DCC in homeostatic regulation of colonic epithelium. This role of DCC is consistent with the concentration gradient of NTN1 along the microvilli axis. The cells in the top portion of the microvilli, where NTN1 levels are low, are in the process of undergoing apoptosis. Conversely,

the cells in the bottom of the microvilli, where NTN1 levels are high, undergo proliferation [22].

However, many other experiments have produced results that questioned whether DCC represents the major tumor suppressor gene in the deleted region of 18q. Firstly, germline mutations that segregate with cancer predisposition are convincing evidence that the candidate gene in question is a tumor suppressor. However, unlike other firmly established tumor suppressors such as *apc* or *rb1*, there is no evidence that DCC germline mutations play a role in a heritable predisposition to colorectal cancer [22]. Furthermore, there are few reports of somatic mutations in DCC in colorectal cancer [22]. The most common somatic mutation is that of the expansion of a dinucleotide repeat tract located in an intron downstream of the seventh exon [24]. However, whether this expansion actually results in decreased expression of DCC has not been demonstrated [22].

1.3.3.2 The SMAD family member 4 (SMAD4) Gene is Another Candidate Tumor Suppressor on 18q

Another candidate tumor suppressor gene in the 18q region is *smad4*. The *smad* genes consist of at least nine members [35], some of which play a critical role in the TGFβ signaling pathway [36]. Binding of the TGFβ ligand to the transmembrane receptor TGFβ-R2 activates it, which results in the phosphorylation of TGFβ-R1. Next, this transmembrane receptor phosphorylates cytoplasmic proteins SMAD2 or SMAD3, which allows the protein to form a hetero-oliogmer with SMAD4 protein. The resulting complex enters the nucleus

to directly or indirectly regulate transcription of target genes, leading to the regulation of cell division.

Germline inactivating mutations of *smad4* result an in increased risk of juvenile polyposis syndrome [37], in which affected individuals develop hamartomatous polyps in the intestine and colon and have an increased risk for colorectal and other gastrointestinal cancers. Furthermore, mice that were heterozygous for *apc* and *smad4* inactivating mutations showed an increased susceptibility of malignant progression compared to mice with only heterozygous inactivating mutations for *apc* [38]. However, because *smad4* is only mutated in about 33% of colorectal cancer samples with chromosome 18q LOH [39-41], it is unlikely to be the major tumor suppressor in the 18q region [22].

1.3.4 p53 Mutation in the Progression of Adenoma to Cancer

Mutations in the *tp53* tumor suppressor gene occur during the transition from adenoma to cancer [4, 42]. The encoded protein, p53, is a transcription factor that functions as a tetramer [43]. The target genes of p53 are involved in many diverse cellular processes such as apoptosis, cell cycle arrest, DNA repair, cellular differentiation, and many others [44]. This 'guardian of the genome' [45] is mutated in approximately 50% of all human cancers [46, 47]. Heterozygous germ-line mutations in *tp53* predispose individuals to a wide range of tumor types at an early age, a condition known as Li-Fraumeni syndrome [48]. Approximately 94% of the tumor-associated mutations of TP53 are point mutations that result in

single amino-acid substitutions [49]. These substitutions can result in a protein that is more stable than wildtype p53 and thus can act as dominant negative inhibitors.

While the studies reported in the literature tend to present the occurrence of these mutations as a linear series of events (Figure 1.1), it should be noted that other studies have shown that it is uncommon for all of the described mutations to be found in one colorectal cancer sample [50, 51], thus emphasizing the heterogeneity of the disease and the critical roles that other genes likely play in cancer initiation and progression.

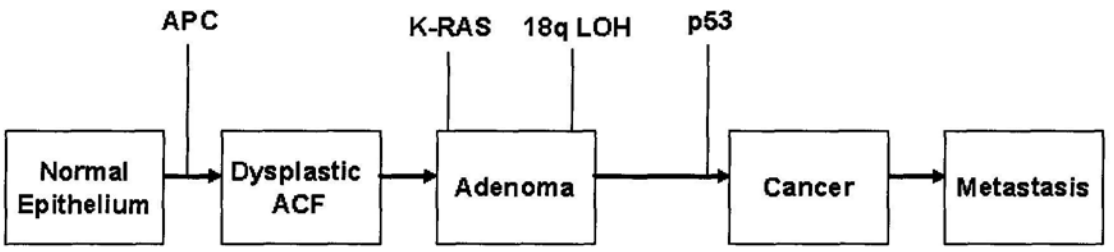

Figure 1.1: Key mutations that occur during the progression of colorectal cancer.

Mutation in K-RAS occurs in the early stages of the adenoma polyp state, while loss of heterozygosity of 18q occurs in the later stages.

1.4 Types of Colorectal Cancer

In general, colorectal cancer is divided into two types: hereditary and non-hereditary (sporadic). Common hereditary colorectal cancer syndromes include

Familial Adenomatous Polyposis (FAP) and Hereditary Nonpolyposis Colorectal Cancer (HNPCC) [52, 53], both of which are due to the occurrence of highly penetrant mutations that result in the development of cancer by approximately 40 years of age, while onset of sporadic colorectal cancer occurs at approximately 67 years of age [5].

1.4.1 Familial Adenomatous Polyposis (FAP)

FAP, an autosomal dominant disease, is caused by mutations that lead to the inactivation of the *apc* gene. Patients with FAP have inherited one mutated *apc* allele from an affected parent and later acquired a somatic mutation of the remaining wildtype *apc* allele. The majority of these mutations are nonsense mutations, thus resulting in a truncated APC protein [17, 52-54]. By age 16, the development of hundreds to thousands of adenoma polyps in the colon has taken place. Without treatment, the average age of cancer onset is 39 years of age [54] with a life expectancy of 42 years [55]. While these individual adenoma polyps are endoscopically and histologically identical to those resulting from sporadic colorectal cancer, their early onset and sheer number guarantees that at least some will acquire further mutations and develop into colorectal cancer [54].

Due to the aggressiveness of the syndrome, early diagnosis and screening are essential. Currently, genetic testing can be used clinically to identify individuals at risk before the development of polyps. Most often, genetic testing involves

DNA sequencing of *apc* to detect mutations followed by complementary methods to test for protein truncation [56, 57]. Once it has been established that an individual is at risk for FAP, screening and surveillance methods are employed to monitor the patient. Flexible sigmoidoscopy screening is performed annually and continues until 35 years of age, at which point the time interval can be reduced to once every three years if no polyps have been observed [58, 59]. For those with FAP, prophylactic colectomy during adolescence is often the treatment [59].

1.4.2 Hereditary Nonpolyposis Colorectal Cancer (HNPCC)

The other major type of hereditary colorectal cancer is HNPCC. Inherited in an autosomal dominant fashion, HNPCC is the most common form of hereditary colorectal cancer [60, 61]. Mutations in five mismatch repair (MMR) genes are responsible for the development of this disorder. In patients with HNPCC, *msh2* and *mlh1* are the most commonly mutated genes, while mutations in *hpms1*, *hpms2*, and *hmsh6* are found less frequently [54]. MMR genes function to recognize and repair DNA replication errors, typically the insertion or deletion (indel) of one or more nucleotides [56, 62]. Defects in MMR genes can often be detected in microsatellites, which are sequences found throughout the genome that contain many one to three nucleotide repeats. Tumor cells in HNPCC patients demonstrate microsatellite instability (MSI), in which these sequences expand or contract in length due to the inability of the cell to repair these DNA replication errors. Therefore, MSI can impact the regular function of a protein. For example, it has been demonstrated that TGFβ-R2, which contains a

polyadenine tract, is responsible for relaying the binding of the TGF-β signal to the nucleus to induce the expression of target genes. This signal transduction pathway has been shown to inhibit the epithelial cell division, thus loss of this receptor has been associated with tumor growth [63, 64]. In many MSI colorectal cancer cell lines, short indels in both alleles results in frameshift mutations, which leads to this gene's inactivation [65-67].

The adenomas produced in HNPCC patients cannot be distinguished endoscopically from those that result from sporadic colon cancer. Adenomas in HNPCC patients develop between 30 and 40 years of age, are larger, and have more villous and dysplastic traits [68, 69] than those in the general population. In the absence of treatment, the average age of cancer onset is 44 years of age [70].

Similar to FAP, genetic tests are in place to identify subjects at risk for HNPCC. To evaluate families at risk for HNPCC, the Bethesda guideline has been developed [57, 62, 71]. This guideline contains of a series of clinical criteria that should be met before genetic testing for HNPCC is conducted. If the guidelines are met, genetic testing is conducted in which five specific DNA microsatellites are evaluated for mutations. If at least two of the five are abnormal, the tumor is labeled as 'MSI-high' [62, 71] and are candidates for *msh2* and *mlh1* genetic testing. If only one of the five is abnormal, the tumor is labeled as 'MSI-low.' Finally, if no abnormalities are found, the tumor is labeled as stable. For patients

at risk of HNPCC, full colonoscopy is recommended once every one to two years beginning at approximately 20 years of age [58, 72] because this has been demonstrated to be effective at reducing incidence of colorectal cancer and mortality [73]. Similar to FAP, prophylactic colectomy is the recommended treatment for HNPCC patients [54].

1.4.3 Sporadic Colorectal Cancer

Data on the movement of immigrants show that people who move from a low-risk area to high-risk area for colorectal cancers rapidly reach equivalent incidence rates to those of the adopted country [74, 75]. This observation suggests that environmental factors play a role in the development of colorectal cancer. Furthermore, hereditary colorectal cancer has been estimated to account for only 20 to 25% of all colorectal cancer cases [53], suggesting that, in fact, the majority of cases of colorectal cancer result from environmental factors.

Diet is one environmental factor that has been well-studied for its association to colorectal cancer incidence. Diets rich in fiber, fruits, and vegetables have been proposed to protect against colorectal cancer [76] and adoption of such a diet could prevent 50 to 75% of colorectal cancer cases [77]. Conversely, high intake of red meat and saturated fats has been shown to have a positive association with the incidence of colorectal cancer by both retrospective [78-83] and prospective [84-86] studies. Furthermore, a recent meta-analysis of prospective studies has demonstrated that consumption of red meat was associated

significantly with increased odds of colorectal cancer [87]. Despite of the abundance of this epidemiological evidence, little is known about the molecular mechanisms that underlie the protective or predisposing effects modulated by diet [88, 89]. However, the molecular mechanisms of initiation and progression of sporadic cases are similar to those described for hereditary cases, with inactivation of *apc* and MMR genes occurring in sporadic cases as well [5].

1.5 Colorectal Cancer Biomarkers

A biomarker is defined as any measurable trait that is indicative of some biological state [90]. Biomarkers exist in many different forms, such as physiological measurements (e.g., weight and blood pressure), images (e.g., mammograms), genetic alterations (e.g., mutations in *apc*), molecules (e.g., prostate-specific antigen, PSA), and others [91]. However, the term 'biomarker' is now more commonly used in the context of measuring molecules from patient samples such as serum, urine, and tissue [90].

Elucidation of biomarkers is an active area of research, as they are valuable in many different aspects of disease management [92]. Firstly, biomarkers could be used in the stratification of patients based on cancer risk to identify those most or least likely to benefit from further screening. For example, women with deleterious mutations in BRCA1 and BRCA2 have markedly increased risk of developing ovarian and breast cancer [93]. Therefore, only the women with these mutations need to undergo aggressive risk reduction strategies, thus

sparing women without these mutations from unnecessary treatment. Secondly, diagnostic biomarkers could be used in tests to screen asymptomatic people for early stages of cancer. By identifying these people, proper treatment can be undertaken earlier, thus potentially resulting in improved prognosis. Prostate-specific antigen (PSA) in serum is an example of such a biomarker used in the early detection and management of prostate cancer [94-96]. If a patient has consistently elevated levels of serum PSA, then further testing, such as with a prostate biopsy, will take place. Thirdly, biomarkers may serve as surrogate endpoints for a therapeutic response that can potentially shorten the length of clinical trials and thus accelerate drug development. For example, a clinical trial is currently comparing four treatment options for chronic myelogenous leukemia (CML) [92]. The success of a treatment will be defined as a greater than four-log reduction in the break point cluster region-Abelson (BCR-ABL) signal. Using the reduction of this DNA marker as the endpoint, instead of the subjects entering remission, will shorten the clinical trial from several years to approximately one year [92]. Finally, biomarkers can also be used to monitor those patients who have gone into remission for cancer recurrence. For example, Genomic Health, a California-based company, demonstrated that its marketed test, Oncotype Dx, which consists of a 21 gene assay, could predict recurrence of tamoxifen-treated, node-negative breast cancer [97].

1.6 Colorectal Cancer Biomarkers Discovery using Transcript based Gene Expression Profiling Methods

Discovery and validation of biomarkers in colorectal cancer can potentially lead to better treatment and management of the disease. For example, biomarkers could be used in a test to diagnose colorectal cancer in asymptomatic patients. Currently, the non-invasive fecal-occult blood test (FOBT) is used to screen patients for colorectal cancer. A stool sample is collected onto a pad that will change colour if the presence of blood is detected in the sample, as determined by peroxidase activity found in hemoglobin [98]. Typically, a colonoscopy is used as the follow-up test to a positive FOBT result. While large randomized clinical trials have shown that FOBT is effective in reducing colorectal cancer mortality by up to 30% [99-101], the test has poor sensitivity in detecting early-stage lesions, as adenomas frequently do not bleed [98]. Furthermore, false positives are also common, as the presence of peroxidase in stool can originate from rare red meat or vegetables in the diet [102]. Thus, highly sensitive and specific diagnostic biomarkers will aid in diagnosing colorectal cancer in subjects.

Another example of biomarkers potentially leading to better treatment and management of colorectal cancer involves patients with Dukes' C colorectal cancer. Post surgical Dukes' C patients have a 60% likelihood of recurrence within five years. Treatment of these patients with post surgical adjuvant therapy reduces the recurrence rate to between 40 and 50% and is thus now the standard treatment for these patients [103-105]. However, because it is currently

not possible to accurately distinguish between those who are surgically cured and those who will experience disease recurrence, the majority of Dukes' C patients are administered adjuvant therapy even if there will be minimal benefits for them [106]. Thus, the discovery of highly sensitive and specific prognostic biomarkers will allow discrimination between patients who need aggressive post surgical adjuvant therapy and those who do not.

With these issues in mind, many researchers have focused on elucidating biomarkers for colorectal cancer by utilizing gene expression profiling methods on tumor and normal colon mucosa samples, which can be obtained during surgery or colonoscopies. Some examples of such studies include the expression profile of colorectal tumors being compared to that of normal mucosa to identify candidate diagnostic biomarkers (Table 1.1, page 22). Other studies have attempted to identify candidate prognostic biomarkers by comparing the expression profiles of tissue samples from patients of one outcome versus those from patients of another, such as those who will experience disease recurrence and those who will not [106, 107].

In a typical study, the differences in gene expression between tissue samples from different biological states are determined to study the underlying molecular differences. Hybridization based methods, such as oligo-nucleotide and cDNA two channel microarrays, involve the hybridization of fluorescently labeled transcripts to slides spotted with nucleic acid probes [108]. Sequence based

methods, such as Serial Analysis of Gene Expression (SAGE), involve sequencing many short tags, each representative of a transcript, from the various tissue types [109]. In both types of methods, statistical and computational methods are applied to the signal, relative hybridization of transcripts in microarrays and relative abundances of tags in SAGE, to determine which genes are differentially expressed between the biological samples under consideration.

1.7 Analysis of Multiple Independent Studies to Identify Biologically Relevant Genes

With the increase of available data from transcript based gene expression profiling experiments, many potentially useful biomarkers have been determined that may have diagnostic and/or prognostic utility in colorectal cancer. However, few reliable biomarkers have resulted in useful tests to be performed in a clinical setting. One explanation for this lack of translational success would be the inconsistency of the results generated by independent studies [1, 89, 110]. High throughput expression profiling studies usually result in tens to thousands of differentially expressed genes, only a small fraction of which are reproduced by independent studies. Explanations for this low overlap include utilization of different tissue resection methods, such as microdissection versus laser capture microdissection, different expression profiling technologies, such as cDNA two-channel microarrays, oligonucleotide microarrays, and SAGE, each with their own inherent biases and limitations [111, 112], as well as different analysis methods such as multiple correction tests and fold change thresholds. Given the low overlap between multiple independent studies, it is difficult to determine

which differentially expressed genes should be further studied and validated as biomarkers.

One logical solution to this problem would be to take the intersection of multiple independent studies to identify the genes that are consistently reported as differentially expressed. It is expected that genes biologically relevant to colorectal cancer tumorigenesis will be the most commonly reported, while those due to the inherent noise or biases in the different experimental methodologies utilized will be underrepresented [113]. Many different groups have determined the overlap between independent expression profiling studies [1, 89, 110]. While such lists are useful, the statistical significance of the overlap is often not considered. One can imagine randomly choosing genes from each expression profiling platform of each independent study, randomly labeling each gene as up- or down-regulated, and observing some overlap due to chance alone. Therefore, it is important to determine which differentially expressed genes were consistently reported in independent colorectal cancer expression profiling studies with a statistically significant frequency because these genes represent the best candidates for further validation as biomarkers with other experimental techniques.

1.8 Meta-Analyses of High Throughput Gene Expression Studies

With the rapid growth of related high throughput gene expression data, efforts have been made to combine this data from indendent studies to increase the power to detect a certain outcome, such as the differential expression of a gene [114]. This process is known as meta-analysis [115]. Many such meta-analysis studies have been conducted on microarray studies related to cancer research. [113, 116-119]. Generally, such meta-analyses involve re-analyzing this raw high throughput gene expression data in a consistent fashion. Thus, due to the fact that raw data is often not made available, most of the cancer specific meta-analyses utilize at most five studies. This limiting factor can potentially affect the number of consistently reported differentially expressed genes.

1.9 Thesis Overview

With these limitations in mind, a novel meta-analysis method was recently developed by our group and applied to published studies of differentially expressed genes in thyroid cancer [120]. This meta-analysis method separated genes demonstrating biological relevance from those that were reported spuriously and has the added advantage of not requiring raw data. The approach involved a vote-counting strategy in which a gene was ranked according to the number of studies reporting its differential expression, the total number of tissue samples utilized in the studies, and the average fold change. That study resulted in a panel of 12 differentially expressed genes reported at a

frequency highly unlikely to have occurred by chance. The panel contained both well known thyroid cancer markers as well as some uncharacterized genes, demonstrating the ability of the meta-analysis method to highlight novel candidate biomarkers. With these results in mind, the objective of the current study was to apply the meta-analysis method to colorectal cancer to determine whether a statistically significant level of overlap between studies could be observed and to identify promising biomarkers. Furthermore, we improved the meta-analysis method by dividing genes into semi-quantitative categories based on the number of tissue samples. The categories were lowest (first quartile, Q1), moderate (interquartile range), and greatest (values greater than those in third quartile, Q3) number of tissue samples. This improvement highlighted genes that may have shown the greatest fold changes, but would have been ranked lower by the original meta-analysis method due to fewer tissue samples studied. We curated published lists of differentially expressed genes from 25 independent studies performing high throughput transcript based gene expression profiling to compare colorectal cancer to normal tissue, adenoma to normal tissue, and colorectal cancer to adenoma tissue. We observed that many genes were consistently reported as differentially expressed in multiple studies and that this overlap was highly significant. The list of candidate biomarkers that resulted from this thesis research will be a valuable resource to the colorectal cancer research community for further studies.

Table 1.1: Twenty-three Colorectal Cancer versus Normal Tissue Expression Profiling Studies Included in Analysis

Ref	Study	Platform	Number of Genes/Features	Up-regulated Features (Mapped)	Down-regulated Features (Mapped)
[121]	Habermann JK et al, 2007	Hs-UniGEM2 human cDNA microarray	9128	24 (23)	34 (29)
[122]	Lin YM et al, 2002	Custom cDNA microarray	23,040	63 (53)	375 (321)
[123]	Buckhaults P et al, 2001	SAGE	N/A	153 (106)	246 (201)
[124]	Notterman DA et al, 2001	Affymetrix Human 6500 GeneChip Set	7457	19 (19)	47 (45)
[125]	Galamb O et al, 2006	Human Atlas Glass 1.0 cDNA microarray	1090	83 (83)	17 (17)
[126]	Wang JY et al, 2005	TGS s-4k cDNA microarray	3800	23 (23)	0
[127]	Croner RS et al, 2005	Affymetrix HG-U133A	22,283	67 (66)	63 (62)
[128]	Kwon HY et al, 2004	Macrogen MAGIC cDNA microarray	4608	77 (77)	45 (44)
[129]	Bertucci F et al, 2004	Custom nylon cDNA microarray	8074	125 (125)	109 (109)
[130]	Ohmachi T et al, 2006	Agilent cDNA microarray	12,814	84 (82)	0
[131]	Mori D et al, 2005	Human Atlas Glass 1.0 cDNA microarray	1090	32 (32)	0
[132]	Kim H et al, 2004	Oligonucleotide microarray from Compugen/Sigma-Genosys	18,861	272 (271)	216 (216)
[133]	Zou TT et al, 2002	Custom cDNA microarray	8000	88 (69)	142 (118)
[134]	Koehler A et al, 2004	Atlas Human Cancer 1.2 Array	1185	31 (29)	14 (13)

Ref	Study	Platform	Number of Genes/Features	Up-regulated Features (Mapped)	Down-regulated Features (Mapped)
[135]	Ichikawa Y et al, 2002	Custom cDNA microarray	20,784	47 (45)	83 (78)
[136]	Jansova E et al, 2006	Human 19K microarrays (Clinical Genomic Centre)	19,201	31 (29)	163 (162)
[137]	Grade M et al, 2007	National Cancer Institute oligonucleotide arrays (Operon V2 oligo set)	21,543	1057 (994)	36 (36)
[138]	Bianchini M et al, 2006	Human 19K microarrays (Clinical Genomic Centre)	19,201	76 (76)	12 (12)
[139]	Agrawal D et al, 2002	Affymetrix Human 6800 GeneChip Set	7129	257 (253)	82 (78)
[140]	Sugiyama Y et al, 2005	Human Cancer Pathway Finder Gene Arrays (Superarray Bioscience)	96	13 (13)	11 (11)
[141]	Kitahara O et al, 2001	Custom cDNA microarray	9216	44 (42)	191 (163)
[142]	Williams NS, et al 2003	Custom cDNA microarray	9592	203 (192)	85 (76)
[143]	Takemasa I et al, 2001	Custom cDNA microarray	4608	22 (22)	36 (36)
Totals				**3582 (3273)**	**2955 (2613)**

Chapter 1 References

1. Shih W, Chetty R, Tsao MS: **Expression profiling by microarrays in colorectal cancer (review).** *Oncol Rep* 2005, **13:**517-524.

2. **Colorectal cancer stats.**
http://www.cancer.ca/ccs/internet/standard/0,2283,3172_14447__langId-en,00.html **Accessed April 2007**

3. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ: **Cancer statistics, 2007.** *CA Cancer J Clin* 2007, **57:**43-66.

4. Fearon ER, Vogelstein B: **A genetic model for colorectal tumorigenesis.** *Cell* 1990, **61:**759-767.

5. Kinzler KW, Vogelstein B: **Lessons from hereditary colorectal cancer.** *Cell* 1996, **87:**159-170.

6. Bird RP: **Role of aberrant crypt foci in understanding the pathogenesis of colon cancer.** *Cancer Lett* 1995, **93:**55-71.

7. Dukes CE: **The classification of cancer of the rectum.** *J Pathol Bacteriol* 1932, **35:**323-332.

8. **Uicc | international union against cancer -** http://www.uicc.org/ **Accessed April 2007**

9. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW: **Identification of c-MYC as a target of the APC pathway.** *Science* 1998, **281:**1509-1512.

10. Tetsu O, McCormick F: **Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells.** *Nature* 1999, **398:**422-426.

11. Brabletz T, Jung A, Dag S, Hlubek F, Kirchner T: **Beta-catenin regulates the expression of the matrix metalloproteinase-7 in human colorectal cancer.** *Am J Pathol* 1999, **155:**1033-1038.

12. Crawford HC, Fingleton BM, Rudolph-Owen LA, Goss KJ, Rubinfeld B, Polakis P, Matrisian LM: **The metalloproteinase matrilysin is a target of beta-catenin transactivation in intestinal tumors.** *Oncogene* 1999, **18:**2883-2891.

13. Schneikert J, Behrens J: **The canonical wnt signalling pathway and its APC partner in colon cancer development.** *Gut* 2007, **56:**417-425.

14. Lengauer C, Kinzler KW, Vogelstein B: **Genetic instability in colorectal cancers.** *Nature* 1997, **386:**623-627.

15. Green RA, Kaplan KB: **Chromosome instability in colorectal tumor cells is associated with defects in microtubule plus-end attachments caused by a dominant mutation in APC.** *J Cell Biol* 2003, **163:**949-961.

16. Green RA, Wollman R, Kaplan KB: **APC and EB1 function together in mitosis to regulate spindle dynamics and chromosome alignment.** *Mol Biol Cell* 2005, **16:**4609-4622.

17. Fodde R, Smits R, Clevers H: **APC, signal transduction and genetic instability in colorectal cancer.** *Nat Rev Cancer* 2001, **1:**55-67.

18. Kaplan KB, Burds AA, Swedlow JR, Bekir SS, Sorger PK, Nathke IS: **A role for the adenomatous polyposis coli protein in chromosome segregation.** *Nat Cell Biol* 2001, **3:**429-432.

19. Bos JL: **Ras oncogenes in human cancer: A review.** *Cancer Res* 1989, **49:**4682-4689.

20. Yamamoto H, Itoh F, Senota A, Adachi Y, Yoshimoto M, Endoh T, Hinoda Y, Yachi A, Imai K: **Expression of matrix metalloproteinase matrilysin (MMP-7) was induced by activated ki-ras via AP-1 activation in SW1417 colon cancer cells.** *J Clin Lab Anal* 1995, **9:**297-301.

21. Smakman N, Borel Rinkes IH, Voest EE, Kranenburg O: **Control of colorectal metastasis formation by K-ras.** *Biochim Biophys Acta* 2005, **1756:**103-114.

22. Mehlen P, Fearon ER: **Role of the dependence receptor DCC in colorectal cancer pathogenesis.** *J Clin Oncol* 2004, **22:**3420-3428.

23. Knudson AG,Jr: **Hereditary cancer, oncogenes, and antioncogenes.** *Cancer Res* 1985, **45:**1437-1443.

24. Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW: **Identification of a chromosome 18q gene that is altered in colorectal cancers.** *Science* 1990, **247:**49-56.

25. Chan SS, Zheng H, Su MW, Wilk R, Killeen MT, Hedgecock EM, Culotti JG: **UNC-40, a C. elegans homolog of DCC (deleted in colorectal cancer), is required in motile cells responding to UNC-6 netrin cues.** *Cell* 1996, **87:**187-195.

26. Keino-Masu K, Masu M, Hinck L, Leonardo ED, Chan SS, Culotti JG, Tessier-Lavigne M: **Deleted in colorectal cancer (DCC) encodes a netrin receptor.** *Cell* 1996, **87:**175-185.

27. Serafini T, Colamarino SA, Leonardo ED, Wang H, Beddington R, Skarnes WC, Tessier-Lavigne M: **Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system.** *Cell* 1996, **87:**1001-1014.

28. Tessier-Lavigne M, Goodman CS: **The molecular biology of axon guidance.** *Science* 1996, **274:**1123-1133.

29. Manitt C, Kennedy TE: **Where the rubber meets the road: Netrin expression and function in developing and adult nervous systems.** *Prog Brain Res* 2002, **137:**425-442.

30. Fazeli A, Dickinson SL, Hermiston ML, Tighe RV, Steen RG, Small CG, Stoeckli ET, Keino-Masu K, Masu M, Rayburn H, Simons J, Bronson RT, Gordon JI, Tessier-Lavigne M, Weinberg RA: **Phenotype of mice lacking functional deleted in colorectal cancer (dcc) gene.** *Nature* 1997, **386:**796-804.

31. Forcet C, Ye X, Granger L, Corset V, Shin H, Bredesen DE, Mehlen P: **The dependence receptor DCC (deleted in colorectal cancer) defines an alternative mechanism for caspase activation.** *Proc Natl Acad Sci U S A* 2001, **98:**3416-3421.

32. Liu J, Yao F, Wu R, Morgan M, Thorburn A, Finley RL,Jr, Chen YQ: **Mediation of the DCC apoptotic signal by DIP13 alpha.** *J Biol Chem* 2002, **277:**26281-26285.

33. Mehlen P, Rabizadeh S, Snipas SJ, Assa-Munt N, Salvesen GS, Bredesen DE: **The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis.** *Nature* 1998, **395:**801-804.

34. Bredesen DE, Ye X, Tasinato A, Sperandio S, Wang JJ, Assa-Munt N, Rabizadeh S: **p75NTR and the concept of cellular dependence: Seeing how the other half die.** *Cell Death Differ* 1998, **5:**365-371.

35. Zhou S, Kinzler KW, Vogelstein B: **Going mad with smads.** *N Engl J Med* 1999, **341:**1144-1146.

36. Massague J, Blain SW, Lo RS: **TGFbeta signaling in growth control, cancer, and heritable disorders.** *Cell* 2000, **103:**295-309.

37. Howe JR, Roth S, Ringold JC, Summers RW, Jarvinen HJ, Sistonen P, Tomlinson IP, Houlston RS, Bevan S, Mitros FA, Stone EM, Aaltonen LA: **Mutations in the SMAD4/DPC4 gene in juvenile polyposis.** *Science* 1998, **280:**1086-1088.

38. Takaku K, Oshima M, Miyoshi H, Matsui M, Seldin MF, Taketo MM: **Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and apc genes.** *Cell* 1998, **92:**645-656.

39. Montgomery E, Goggins M, Zhou S, Argani P, Wilentz R, Kaushal M, Booker S, Romans K, Bhargava P, Hruban R, Kern S: **Nuclear localization of Dpc4 (Madh4, Smad4) in colorectal carcinomas and relation to mismatch repair/transforming growth factor-beta receptor defects.** *Am J Pathol* 2001, **158:**537-542.

40. Korchynskyi O, Landstrom M, Stoika R, Funa K, Heldin CH, ten Dijke P, Souchelnytskyi S: **Expression of smad proteins in human colorectal cancer.** *Int J Cancer* 1999, **82:**197-202.

41. Salovaara R, Roth S, Loukola A, Launonen V, Sistonen P, Avizienyte E, Kristo P, Jarvinen H, Souchelnytskyi S, Sarlomo-Rikala M, Aaltonen LA: **Frequent loss of SMAD4/DPC4 protein in colorectal cancers.** *Gut* 2002, **51:**56-59.

42. Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D: **Identification of FAP locus genes from chromosome 5q21.** *Science* 1991, **253:**661-665.

43. de Vries A, Flores ER, Miranda B, Hsieh HM, van Oostrom CT, Sage J, Jacks T: **Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function.** *Proc Natl Acad Sci U S A* 2002, **99:**2948-2953.

44. Levine AJ: **P53, the cellular gatekeeper for growth and division.** *Cell* 1997, **88:**323-331.

45. Stiewe T: **The p53 family in differentiation and tumorigenesis.** *Nat Rev Cancer* 2007, **7:**165-168.

46. Greenblatt MS, Bennett WP, Hollstein M, Harris CC: **Mutations in the p53 tumor suppressor gene: Clues to cancer etiology and molecular pathogenesis.** *Cancer Res* 1994, **54:**4855-4878.

47. Hollstein M, Sidransky D, Vogelstein B, Harris CC: **P53 mutations in human cancers.** *Science* 1991, **253:**49-53.

48. Malkin D: **p53 and the li-fraumeni syndrome.** *Biochim Biophys Acta* 1994, **1198:**197-213.

49. Vousden KH, Lu X: **Live or let die: The cell's response to p53.** *Nat Rev Cancer* 2002, **2:**594-604.

50. Smith G, Carey FA, Beattie J, Wilkie MJ, Lightfoot TJ, Coxhead J, Garner RC, Steele RJ, Wolf CR: **Mutations in APC, kirsten-ras, and p53--alternative genetic pathways to colorectal cancer.** *Proc Natl Acad Sci U S A* 2002, **99:**9433-9438.

51. Conlin A, Smith G, Carey FA, Wolf CR, Steele RJ: **The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma.** *Gut* 2005, **54:**1283-1286.

52. Galiatsatos P, Foulkes WD: **Familial adenomatous polyposis.** *Am J Gastroenterol* 2006, **101:**385-398.

53. de la Chapelle A: **Genetic predisposition to colorectal cancer.** *Nat Rev Cancer* 2004, **4:**769-780.

54. Strate LL, Syngal S: **Hereditary colorectal cancer syndromes.** *Cancer Causes Control* 2005, **16:**201-213.

55. Galle TS, Juel K, Bulow S: **Causes of death in familial adenomatous polyposis.** *Scand J Gastroenterol* 1999, **34:**808-812.

56. Giardiello FM, Brensinger JD, Petersen GM: **AGA technical review on hereditary colorectal cancer and genetic testing.** *Gastroenterology* 2001, **121:**198-213.

57. Kaz AM, Brentnall TA: **Genetic testing for colon cancer.** *Nat Clin Pract Gastroenterol Hepatol* 2006, **3:**670-679.

58. Winawer SJ, Fletcher RH, Miller L, Godlee F, Stolar MH, Mulrow CD, Woolf SH, Glick SN, Ganiats TG, Bond JH, Rosen L, Zapka JG, Olsen SJ, Giardiello FM, Sisk JE, Van Antwerp R, Brown-Davis C, Marciniak DA, Mayer RJ: **Colorectal cancer screening: Clinical guidelines and rationale.** *Gastroenterology* 1997, **112:**594-642.

59. King JE, Dozois RR, Lindor NM, Ahlquist DA: **Care of patients and their families with familial adenomatous polyposis.** *Mayo Clin Proc* 2000, **75:**57-67.

60. Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P, Chadwick RB, Kaariainen H, Eskelinen M, Jarvinen H, Mecklin JP, de la Chapelle A: **Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease.** *N Engl J Med* 1998, **338:**1481-1487.

61. Samowitz WS, Curtin K, Lin HH, Robertson MA, Schaffer D, Nichols M, Gruenthal K, Leppert MF, Slattery ML: **The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer.** *Gastroenterology* 2001, **121:**830-838.

62. Burt R, Neklason DW: **Genetic testing for inherited colon cancer.** *Gastroenterology* 2005, **128:**1696-1716.

63. Bierie B, Moses HL: **Tumour microenvironment: TGFbeta: The molecular jekyll and hyde of cancer.** *Nat Rev Cancer* 2006, **6:**506-520.

64. Pardali K, Moustakas A: **Actions of TGF-beta as tumor suppressor and pro-metastatic factor in human cancer.** *Biochim Biophys Acta* 2007, **1775:**21-62.

65. Bacon AL, Farrington SM, Dunlop MG: **Mutation frequency in coding and non-coding repeat sequences in mismatch repair deficient cells derived from normal human tissue.** *Oncogene* 2001, **20:**7464-7471.

66. Parsons R, Myeroff LL, Liu B, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B: **Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer.** *Cancer Res* 1995, **55:**5548-5550.

67. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B: **Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability.** *Science* 1995, **268:**1336-1338.

68. Lynch HT, Watson P, Lanspa SJ, Marcus J, Smyrk T, Fitzgibbons RJ,Jr, Kriegler M, Lynch JF: **Natural history of colorectal cancer in hereditary nonpolyposis colorectal cancer (lynch syndromes I and II).** *Dis Colon Rectum* 1988, **31:**439-444.

69. Vasen HF, Mecklin JP, Watson P, Utsunomiya J, Bertario L, Lynch P, Svendsen LB, Cristofaro G, Muller H, Khan PM: **Surveillance in hereditary nonpolyposis colorectal cancer: An international cooperative study of 165 families. the international collaborative group on HNPCC.** *Dis Colon Rectum* 1993, **36:**1-4.

70. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomaki P, Mecklin JP, Jarvinen HJ: **Cancer risk in mutation carriers of DNA-mismatch-repair genes.** *Int J Cancer* 1999, **81:**214-218.

71. Grady WM: **Genetic testing for high-risk colon cancer patients.** *Gastroenterology* 2003, **124:**1574-1594.

72. Burke W, Petersen G, Lynch P, Botkin J, Daly M, Garber J, Kahn MJ, McTiernan A, Offit K, Thomson E, Varricchio C: **Recommendations for follow-up care of individuals with an inherited predisposition to cancer. I. hereditary nonpolyposis colon cancer. cancer genetics studies consortium.** *JAMA* 1997, **277:**915-919.

73. Jarvinen HJ, Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomaki P, De La Chapelle A, Mecklin JP: **Controlled 15-year trial on screening for**

colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2000, **118:**829-834.

74. Haenszel W, Kurihara M: **Studies of japanese migrants. I. mortality from cancer and other diseases among japanese in the united states.** *J Natl Cancer Inst* 1968, **40:**43-68.

75. McMichael AJ, Giles GG: **Cancer in migrants to australia: Extending the descriptive epidemiological data.** *Cancer Res* 1988, **48:**751-756.

76. Bingham SA, Day NE, Luben R, Ferrari P, Slimani N, Norat T, Clavel-Chapelon F, Kesse E, Nieters A, Boeing H, Tjonneland A, Overvad K, Martinez C, Dorronsoro M, Gonzalez CA, Key TJ, Trichopoulou A, Naska A, Vineis P, Tumino R, Krogh V, Bueno-de-Mesquita HB, Peeters PH, Berglund G, Hallmans G, Lund E, Skeie G, Kaaks R, Riboli E, European Prospective Investigation into Cancer and Nutrition: **Dietary fibre in food and protection against colorectal cancer in the european prospective investigation into cancer and nutrition (EPIC): An observational study.** *Lancet* 2003, **361:**1496-1501.

77. Doll R, Peto R: **The causes of cancer: Quantitative estimates of avoidable risks of cancer in the united states today.** *J Natl Cancer Inst* 1981, **66:**1191-1308.

78. Jain M, Cook GM, Davis FG, Grace MG, Howe GR, Miller AB: **A case-control study of diet and colo-rectal cancer.** *Int J Cancer* 1980, **26:**757-768.

79. Potter JD, McMichael AJ: **Diet and cancer of the colon and rectum: A case-control study.** *J Natl Cancer Inst* 1986, **76:**557-569.

80. Lyon JL, Mahoney AW, West DW, Gardner JW, Smith KR, Sorenson AW, Stanish W: **Energy intake: Its relationship to colon cancer risk.** *J Natl Cancer Inst* 1987, **78:**853-861.

81. Whittemore AS, Wu-Williams AH, Lee M, Zheng S, Gallagher RP, Jiao DA, Zhou L, Wang XH, Chen K, Jung D: **Diet, physical activity, and colorectal cancer among chinese in north america and china.** *J Natl Cancer Inst* 1990, **82:**915-926.

82. Miller AB, Howe GR, Jain M, Craib KJ, Harrison L: **Food items and food groups as risk factors in a case-control study of diet and colo-rectal cancer.** *Int J Cancer* 1983, **32:**155-161.

83. Gerhardsson de Verdier M, Hagman U, Peters RK, Steineck G, Overvik E: **Meat, cooking methods and colorectal cancer: A case-referent study in stockholm.** *Int J Cancer* 1991, **49:**520-525.

84. Goldbohm RA, van den Brandt PA, van 't Veer P, Brants HA, Dorant E, Sturmans F, Hermus RJ: **A prospective cohort study on the relation between**

meat consumption and the risk of colon cancer. *Cancer Res* 1994, **54:**718-723.

85. English DR, MacInnis RJ, Hodge AM, Hopper JL, Haydon AM, Giles GG: **Red meat, chicken, and fish consumption and risk of colorectal cancer.** *Cancer Epidemiol Biomarkers Prev* 2004, **13:**1509-1514.

86. Chao A, Thun MJ, Connell CJ, McCullough ML, Jacobs EJ, Flanders WD, Rodriguez C, Sinha R, Calle EE: **Meat consumption and risk of colorectal cancer.** *JAMA* 2005, **293:**172-182.

87. Larsson SC, Wolk A: **Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies.** *Int J Cancer* 2006, **119:**2657-2664.

88. Jacobs ET, Thompson PA, Martinez ME: **Diet, gender, and colorectal neoplasia.** *J Clin Gastroenterol* 2007, **41:**731-746.

89. Cardoso J, Boer J, Morreau H, Fodde R: **Expression and genomic profiling of colorectal cancer.** *Biochim Biophys Acta* 2007, **1775:**103-137.

90. Rifai N, Gillette MA, Carr SA: **Protein biomarker discovery and validation: The long and uncertain path to clinical utility.** *Nat Biotechnol* 2006, **24:**971-983.

91. Dalton WS, Friend SH: **Cancer biomarkers--an invitation to the table.** *Science* 2006, **312:**1165-1168.

92. Hartwell L, Mankoff D, Paulovich A, Ramsey S, Swisher E: **Cancer biomarkers: A systems approach.** *Nat Biotechnol* 2006, **24:**905-908.

93. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M: **Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. the breast cancer linkage consortium.** *Am J Hum Genet* 1998, **62:**676-689.

94. Catalona WJ, Richie JP, deKernion JB, Ahmann FR, Ratliff TL, Dalkin BL, Kavoussi LR, MacFarlane MT, Southwick PC: **Comparison of prostate specific antigen concentration versus prostate specific antigen density in the early detection of prostate cancer: Receiver operating characteristic curves.** *J Urol* 1994, **152:**2031-2036.

95. Leach FS, Koh MS, Chan YW, Bark S, Ray R, Morton RA, Remaley AT: **Prostate specific antigen as a clinical biomarker for prostate cancer: What's the take home message?** *Cancer Biol Ther* 2005, **4:**371-375.

96. Brawer MK, Chetner MP, Beatie J, Buchner DM, Vessella RL, Lange PH: **Screening for prostatic carcinoma with prostate specific antigen.** *J Urol* 1992, **147:**841-845.

97. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351:**2817-2826.

98. Booth RA: **Minimally invasive biomarkers for detection and staging of colorectal cancer.** *Cancer Lett* 2007, **249:**87-96.

99. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, Ederer F: **Reducing mortality from colorectal cancer by screening for fecal occult blood. minnesota colon cancer control study.** *N Engl J Med* 1993, **328:**1365-1371.

100. Mandel JS, Church TR, Ederer F, Bond JH: **Colorectal cancer mortality: Effectiveness of biennial screening for fecal occult blood.** *J Natl Cancer Inst* 1999, **91:**434-437.

101. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, James PD, Mangham CM: **Randomised controlled trial of faecal-occult-blood screening for colorectal cancer.** *Lancet* 1996, **348:**1472-1477.

102. Walsh JM, Terdiman JP: **Colorectal cancer screening: Scientific review.** *JAMA* 2003, **289:**1288-1296.

103. Saltz LB, Kelsen DP: **Adjuvant treatment of colorectal cancer.** *Annu Rev Med* 1997, **48:**191-202.

104. Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Tangen CM, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, Veeder MH, Mailliard JA: **Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: A final report.** *Ann Intern Med* 1995, **122:**321-326.

105. Wolmark N, Rockette H, Mamounas E, Jones J, Wieand S, Wickerham DL, Bear HD, Atkins JN, Dimitrov NV, Glass AG, Fisher ER, Fisher B: **Clinical trial to assess the relative efficacy of fluorouracil and leucovorin, fluorouracil and levamisole, and fluorouracil, leucovorin, and levamisole in patients with dukes' B and C carcinoma of the colon: Results from national surgical adjuvant breast and bowel project C-04.** *J Clin Oncol* 1999, **17:**3553-3559.

106. Arango D, Laiho P, Kokko A, Alhopuro P, Sammalkorpi H, Salovaara R, Nicorici D, Hautaniemi S, Alazzouzi H, Mecklin JP, Jarvinen H, Hemminki A, Astola J, Schwartz S,Jr, Aaltonen LA: **Gene-expression profiling predicts recurrence in dukes' C colorectal cancer.** *Gastroenterology* 2005, **129:**874-884.

107. Wang Y, Jatkoe T, Zhang Y, Mutch MG, Talantov D, Jiang J, McLeod HL, Atkins D: **Gene expression profiles and molecular markers to predict recurrence of dukes' B colon cancer.** *J Clin Oncol* 2004, **22:**1564-1571.

108. Cowell JK, Hawthorn L: **The application of microarray technology to the analysis of the cancer genome.** *Curr Mol Med* 2007, **7:**103-120.

109. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270:**484-487.

110. Sagynaliev E, Steinert R, Nestler G, Lippert H, Knoch M, Reymond MA: **Web-based data warehouse on gene expression in human colorectal cancer.** *Proteomics* 2005, **5:**3066-3078.

111. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS: **Analysis of matched mRNA measurements from two different microarray technologies.** *Bioinformatics* 2002, **18:**405-412.

112. Siddiqui AS, Delaney AD, Schnerch A, Griffith OL, Jones SJ, Marra MA: **Sequence biases in large scale gene expression profiling data.** *Nucleic Acids Res* 2006, **34:**e83.

113. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression.** *Proc Natl Acad Sci U S A* 2004, **101:**9309-9314.

114. Ghosh D, Barette TR, Rhodes D, Chinnaiyan AM: **Statistical issues and methods for meta-analysis of microarray data: A case study in prostate cancer.** *Funct Integr Genomics* 2003, **3:**180-188.

115. Normand SL: **Meta-analysis: Formulating, evaluating, combining, and reporting.** *Stat Med* 1999, **18:**321-359.

116. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM: **Meta-analysis of microarrays: Interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer.** *Cancer Res* 2002, **62:**4427-4433.

117. Rhodes DR, Chinnaiyan AM: **Integrative analysis of the cancer transcriptome.** *Nat Genet* 2005, **37 Suppl:**S31-7.

118. Jiang H, Deng Y, Chen HS, Tao L, Sha Q, Chen J, Tsai CJ, Zhang S: **Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes.** *BMC Bioinformatics* 2004, **5:**81.

119. Yang X, Sun X: **Meta-analysis of several gene lists for distinct types of cancer: A simple way to reveal common prognostic markers.** *BMC Bioinformatics* 2007, **8:**118.

120. Griffith OL, Melck A, Jones SJ, Wiseman SM: **Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers.** *J Clin Oncol* 2006, **24:**5043-5051.

121. Habermann JK, Paulsen U, Roblick UJ, Upender MB, McShane LM, Korn EL, Wangsa D, Kruger S, Duchrow M, Bruch HP, Auer G, Ried T: **Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis.** *Genes Chromosomes Cancer* 2007, **46:**10-26.

122. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y: **Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas.** *Oncogene* 2002, **21:**4120-4128.

123. Buckhaults P, Rago C, St Croix B, Romans KE, Saha S, Zhang L, Vogelstein B, Kinzler KW: **Secreted and cell surface genes expressed in benign and malignant colorectal tumors.** *Cancer Res* 2001, **61:**6996-7001.

124. Notterman DA, Alon U, Sierk AJ, Levine AJ: **Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays.** *Cancer Res* 2001, **61:**3124-3130.

125. Galamb O, Sipos F, Dinya E, Spisak S, Tulassay Z, Molnar B: **mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray.** *World J Gastroenterol* 2006, **12:**6998-7006.

126. Wang JY, Yeh CS, Tzou WS, Hsieh JS, Chen FM, Lu CY, Yu FJ, Cheng TL, Huang TJ, Lin SR: **Analysis of progressively overexpressed genes in tumorigenesis of colorectal cancers using cDNA microarray.** *Oncol Rep* 2005, **14:**65-72.

127. Croner RS, Foertsch T, Brueckl WM, Guenther K, Siebenhaar R, Stremmel C, Matzel KE, Papadopoulos T, Kirchner T, Behrens J, Klein-Hitpass L, Stuerzl M, Hohenberger W, Reingruber B: **Common denominator genes that distinguish colorectal carcinoma from normal mucosa.** *Int J Colorectal Dis* 2005, **20:**353-362.

128. Kwon HC, Kim SH, Roh MS, Kim JS, Lee HS, Choi HJ, Jeong JS, Kim HJ, Hwang TH: **Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer.** *Dis Colon Rectum* 2004, **47:**141-152.

129. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loriod B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V,

Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D, Houlgatte R: **Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters.** *Oncogene* 2004, **23:**1377-1391.

130. Ohmachi T, Tanaka F, Mimori K, Inoue H, Yanaga K, Mori M: **Clinical significance of TROP2 expression in colorectal cancer.** *Clin Cancer Res* 2006, **12:**3057-3063.

131. Mori D, Nakafusa Y, Miyazaki K, Tokunaga O: **Differential expression of janus kinase 3 (JAK3), matrix metalloproteinase 13 (MMP13), heat shock protein 60 (HSP60), and mouse double minute 2 (MDM2) in human colorectal cancer progression using human cancer cDNA microarrays.** *Pathol Res Pract* 2005, **201:**777-789.

132. Kim H, Nam SW, Rhee H, Shan Li L, Ju Kang H, Hye Koh K, Kyu Kim N, Song J, Tak-Bun Liu E, Kim H: **Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas.** *Oncogene* 2004, **23:**6218-6225.

133. Zou TT, Selaru FM, Xu Y, Shustova V, Yin J, Mori Y, Shibata D, Sato F, Wang S, Olaru A, Deacu E, Liu TC, Abraham JM, Meltzer SJ: **Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon.** *Oncogene* 2002, **21:**4855-4862.

134. Koehler A, Bataille F, Schmid C, Ruemmele P, Waldeck A, Blaszyk H, Hartmann A, Hofstaedter F, Dietmaier W: **Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage.** *J Pathol* 2004, **204:**65-74.

135. Ichikawa Y, Ishikawa T, Takahashi S, Hamaguchi Y, Morita T, Nishizuka I, Yamaguchi S, Endo I, Ike H, Togo S, Oki S, Shimada H, Kadota K, Nakamura S, Goto H, Nitanda H, Satomi S, Sakai T, Narita I, Gejyo F, Tomaru Y, Shimizu K, Hayashizaki Y, Okazaki Y: **Identification of genes regulating colorectal carcinogenesis by using the algorithm for diagnosing malignant state method.** *Biochem Biophys Res Commun* 2002, **296:**497-506.

136. Jansova E, Koutna I, Krontorad P, Svoboda Z, Krivankova S, Zaloudik J, Kozubek M, Kozubek S: **Comparative transcriptome maps: A new approach to the diagnosis of colorectal carcinoma patients using cDNA microarrays.** *Clin Genet* 2006, **69:**218-227.

137. Grade M, Hormann P, Becker S, Hummon AB, Wangsa D, Varma S, Simon R, Liersch T, Becker H, Difilippantonio MJ, Ghadimi BM, Ried T: **Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas.** *Cancer Res* 2007, **67:**41-56.

138. Bianchini M, Levy E, Zucchini C, Pinski V, Macagno C, De Sanctis P, Valvassori L, Carinci P, Mordoh J: **Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa.** *Int J Oncol* 2006, **29:**83-94.

139. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ: **Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling.** *J Natl Cancer Inst* 2002, **94:**513-521.

140. Sugiyama Y, Farrow B, Murillo C, Li J, Watanabe H, Sugiyama K, Evers BM: **Analysis of differential gene expression patterns in colon cancer and cancer stroma using microdissected tissues.** *Gastroenterology* 2005, **128:**480-486.

141. Kitahara O, Furukawa Y, Tanaka T, Kihara C, Ono K, Yanagawa R, Nita ME, Takagi T, Nakamura Y, Tsunoda T: **Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia.** *Cancer Res* 2001, **61:**3544-3549.

142. Williams NS, Gaynor RB, Scoggin S, Verma U, Gokaslan T, Simmang C, Fleming J, Tavana D, Frenkel E, Becerra C: **Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference.** *Clin Cancer Res* 2003, **9:**931-946.

143. Takemasa I, Higuchi H, Yamamoto H, Sekimoto M, Tomita N, Nakamori S, Matoba R, Monden M, Matsubara K: **Construction of preferential cDNA microarray specialized for human colorectal carcinoma: Molecular sketch of colorectal cancer.** *Biochem Biophys Res Commun* 2001, **285:**1244-1249.

Chapter 2: Materials and Methods

2.1 Meta-analysis of Published Datasets of Differentially Expressed Genes in Colorectal Cancer

The meta-analysis method that was created for this thesis was based on a method that was previously developed by our group [1]. This previous method utilized a vote-counting strategy in which a gene was ranked based on three criteria. Presented in order of importance, the three criteria were: 1) the number of studies reporting its differential expression, 2) the total number of tissue samples utilized in the studies, and 3) the average fold change. Originally, the number of tissue samples utilized was deemed more important than the average fold change because many studies do not report a fold change. However, one limitation of this method was that certain genes, while showing a large fold change, were ranked lower due to fewer tissue samples included in the studies. To overcome this, we improved the meta-analysis method by dividing genes into semi-quantitative categories based on the number of tissue samples instead of the absolute numbers. Each gene that was reported in an equal number of independent studies was divided into three semi-quantitative categories based on the number of tissue samples: lowest (first quartile, Q1), moderate (interquartile range), and greatest (values greater than those in third quartile, Q3). This enabled the current version of the meta-analysis method to further emphasize the genes that show the greatest fold changes, which is likely of greater interest.

2.1.1 Data Collection and Curation

NCBI PubMed was queried for high throughput colorectal cancer expression profiling studies between 2000 and 2007. Only studies utilizing human tissue samples obtained from the surgical resection of cancerous tumors and/or adenomatous polyps were considered. Studies were divided into three comparison types: cancer versus normal, adenoma versus normal, and cancer versus adenoma. Studies that focused on determining differentially expressed genes between tissues of varying microsatellite stability, specific stages of colorectal cancer, or those comparing cancer to cancer samples to determine prognostic biomarkers were excluded. In total, differentially expressed genes from 25 independent studies were collected. Twenty-three studies performed expression profiling to compare cancer versus normal samples (**Table 1.1**, page 22), while seven and five studies considered adenoma versus normal (Table 2.1, page 41), and cancer versus adenoma (Table 2.2, page 42), respectively.

2.1.2 Differentially Expressed Genes Mapped to Entrez Gene IDs

In the microarray expression profiling studies, differentially expressed genes were represented by a GenBank accession ID [2], HUGO gene name [3], or Affymetrix probe id [4]. The sequence identifier was mapped to the NCBI Entrez Gene Identifier (Entrez Gene ID) [5] with the aid of custom developed Perl scripts and the Clone/Gene ID Converter tool [6]. For the SAGE study, updated tag to gene mapping data were obtained from SAGE Genie [7].

2.1.3 Total Gene Lists for Each Study

In order to estimate the amount of overlap occurring between studies by chance, we obtained the platform-specific annotation file to identify the genes that could potentially be detected as differentially expressed. For commercial platforms, such as Affymetrix and Atlas microarrays, the annotation file was obtained directly from the company website. The identifiers in these annotation files were mapped to the corresponding Entrez Gene ID as above to produce a total gene list for each study. Identifiers that could not be mapped to an Entrez Gene ID were ignored. To obtain a total gene list for the SAGE study, all gene names in the tag to gene mapping data from SAGE Genie were mapped to Entrez Gene IDs. For studies that utilized platforms in which an annotation file could not be obtained, such as the custom cDNA microarrays and some of the oligonucletotide microarrays, an approximation approach was used in which the appropriate number of Entrez Gene IDs was randomly chosen from the combined gene lists from the other platforms. For example, if a study reported 200 differentially expressed genes, then 200 Entrez Gene IDs were randomly chosen from the combined gene list.

2.1.4 Assessment of Significance of Study Overlap using Simulations

To assess the statistical significance of the observed overlap, custom Perl (version 5.8.3) scripts were created to perform Monte Carlo simulations. In each of the 10,000 permutations, the appropriate number of Entrez Gene IDs from the total gene list of each study was randomly chosen. Next, the appropriate number

of genes were labeled as 'UP' for up-regulated or 'DOWN' for down-regulated. For example, if a study reported 50 up-regulated and 20 down-regulated genes, 70 Entrez Gene IDs were randomly chosen and 50 would be labeled as 'UP' and 20 would be labeled as 'DOWN.' We utilized an 'all-or-none' approach, in which the level of overlap for a particular gene was only considered if all the independent studies reporting its differential expression agreed on the direction of differential expression. The level of overlap between studies in each permutation was counted as in the real analysis. Upon completion of the permutations, a distribution of overlap results from the simulations was determined and a P-value estimated by comparing the number of instances of a certain level of overlap from the simulations to the actual level of overlap in the real data. Significance was defined at $P < 0.05$. A pipeline summarizing the steps undertaken is shown in Figure 2.1.

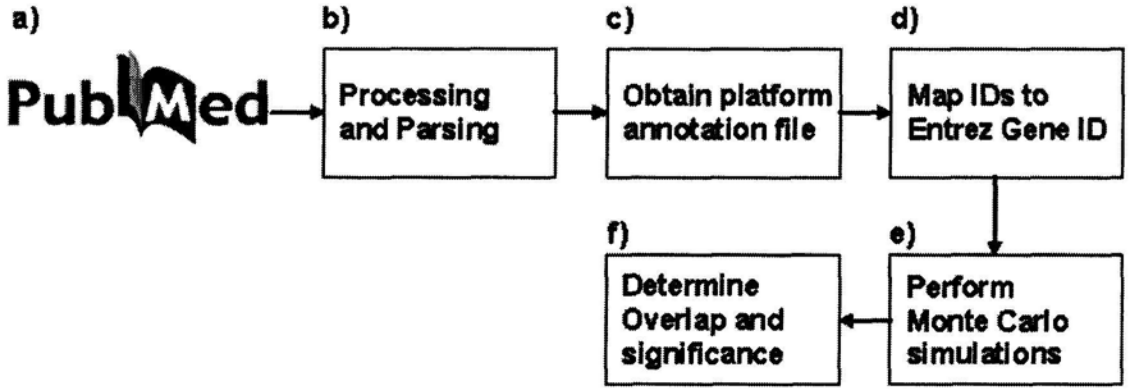

Figure 2.1: A pipeline summarizing the steps undertaken.

a) PubMed was queried to obtain published studies performing transcript based expression profiling to compare cancer versus normal, adenoma versus normal, and cancer versus adenoma. b) Custom Perl scripts were developed to process and parse the lists of differentially expressed genes. c) The platform specific annotation file for each study was obtained. d) The sequence IDs for the list of differentially expressed genes and the platform specific annotation file were mapped to Entrez Gene IDs. e) Monte Carlo simulations were conducted using custom Perl scripts. f) The observed overlap was compared to that of the simulations and statistical significance was assessed.

Table 2.1: Seven Colorectal Adenoma versus Normal Tissue Expression Profiling Studies Included in Analysis

Ref	Study	Platform	Number of Genes/ Features	Up-regulated Genes/ Features (Mapped)	Down-regulated Genes/ Features (Mapped)
[8]	Habermann JK et al, 2007	Hs-UniGEM2 human cDNA microarray	9128	20 (19)	38 (35)
[9]	Lin YM et al, 2002	Custom cDNA microarray	23,040	63 (53)	375 (321)
[10]	Buckhaults P et al, 2001	SAGE	N/A	247 (208)	246 (180)
[11]	Notterman DA et al, 2001	Affymetrix Human 6800 GeneChip Set	7129	20 (20)	0
[12]	Galamb O et al, 2006	Human Atlas Glass 1.0 cDNA microarray	1090	12 (12)	33 (33)
[13]	Wang JY et al, 2005	TGS s-4k cDNA microarray	3800	23 (23)	0
[14]	Lechner S et al, 2003	Atlas Human Cancer cDNA microarray	588	15 (11)	9 (5)
Totals				**400 (346)**	**701 (640)**

Table 2.2: Five Cancer versus Adenoma Tissue Expression Profiling Studies Included in Analysis

Ref	Study	Platform	Number of Genes/ Features	Up-regulated Genes/ Features (Mapped)	Down-regulated Genes/ Features (Mapped)
[8]	Habermann JK et al, 2007	Hs-UniGEM2 human cDNA microarray	9128	80 (80)	36 (32)
[9]	Lin YM et al, 2002	Custom cDNA microarray	23,040	18 (18)	32 (31)
[10]	Buckhaults P et al, 2001	SAGE	N/A	163 (75)	167 (137)
[11]	Notterman DA et al, 2001	Affymetrix Human 6500 GeneChip Set	7457	20 (20)	0
[15]	Nosho K et al, 2005	Gene Navigator cDNA array filter (Toyobo)	561	10 (10)	12 (12)
Totals				**291 (203)**	**247 (212)**

Chapter 2 References

1. Griffith OL, Melck A, Jones SJ, Wiseman SM: **Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers.** *J Clin Oncol* 2006, **24:**5043-5051.

2. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2007, **35:**D21-5.

3. Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA, Lush MJ: **The HUGO gene nomenclature database, 2006 updates.** *Nucleic Acids Res* 2006, **34:**D319-21.

4. Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG: **The affymetrix GeneChip platform: An overview.** *Methods Enzymol* 2006, **410:**3-28.

5. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez gene: Gene-centered information at NCBI.** *Nucleic Acids Res* 2007, **35:**D26-31.

6. Alibes A, Yankilevich P, Canada A, Diaz-Uriarte R: **IDconverter and IDClight: Conversion and annotation of gene and protein IDs.** *BMC Bioinformatics* 2007, **8:**9.

7. Boon K, Osorio EC, Greenhut SF, Schaefer CF, Shoemaker J, Polyak K, Morin PJ, Buetow KH, Strausberg RL, De Souza SJ, Riggins GJ: **An anatomy of normal and malignant gene expression.** *Proc Natl Acad Sci U S A* 2002, **99:**11287-11292.

8. Habermann JK, Paulsen U, Roblick UJ, Upender MB, McShane LM, Korn EL, Wangsa D, Kruger S, Duchrow M, Bruch HP, Auer G, Ried T: **Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis.** *Genes Chromosomes Cancer* 2007, **46:**10-26.

9. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y: **Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas.** *Oncogene* 2002, **21:**4120-4128.

10. Buckhaults P, Rago C, St Croix B, Romans KE, Saha S, Zhang L, Vogelstein B, Kinzler KW: **Secreted and cell surface genes expressed in benign and malignant colorectal tumors.** *Cancer Res* 2001, **61:**6996-7001.

11. Notterman DA, Alon U, Sierk AJ, Levine AJ: **Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays.** *Cancer Res* 2001, **61:**3124-3130.

12. Galamb O, Sipos F, Dinya E, Spisak S, Tulassay Z, Molnar B: **mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray.** *World J Gastroenterol* 2006, **12:**6998-7006.

13. Wang JY, Yeh CS, Tzou WS, Hsieh JS, Chen FM, Lu CY, Yu FJ, Cheng TL, Huang TJ, Lin SR: **Analysis of progressively overexpressed genes in tumorigenesis of colorectal cancers using cDNA microarray.** *Oncol Rep* 2005, **14:**65-72.

14. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F: **Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia.** *Gut* 2003, **52:**1148-1153.

15. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K: **Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis.** *Br J Cancer* 2005, **92:**1193-1200.

Chapter 3: Results

3.1 Mapping Success Rate and Significance of Overlap Observed

Of the total 8176 differentially expressed genes reported in the 25 studies (4273 up- and 3903 down-regulated), 7287 (89.1%) could be mapped to an Entrez Gene ID (3822 up- and 3465 down-regulated). In the cancer versus normal and adenoma versus normal comparisons, significant overlap was observed. No such significance in overlap was observed in the cancer versus adenoma comparison (Table 3.1, page 49), even though each individual study identified differentially expressed genes.

3.2 Overlap Results for Cancer versus Normal Tissue Comparisons

The simulations demonstrated that the amount of overlap in the cancer versus normal tissue comparison was highly significant ($P < .0001$), with 573 genes reported as differentially expressed with consistent direction of change in at last two studies (multi-study genes) (Figure 3.1). There were 175 multi-study genes that were reported with inconsistent direction of differential expression. Thus, the majority of multi-study genes (76.6%) that were reported as differentially expressed agreed on the direction, even for large numbers of studies.

From the cancer versus normal Monte Carlo simulations, an average of 258.30 (95% CI, 258.16 to 258.45) genes would be expected to have an overlap of two, while the actual data contained 410. An average of 18.37 (95% CI, 18.33 to 18.42) genes would be expected to have an overlap of three, compared with 95 in the actual data. For an overlap of four, the simulation produced 1.14 (95% CI, 1.13 to 1.15) genes, while the actual data contained 30 genes. Overlaps of five, six, and seven were observed in the simulations, but with averages of less than one hundredth of a gene. In 10,000 permutations, the simulations never produced an overlap greater than seven, while two genes had an overlap of nine and one gene had an overlap of 11 in the real data. In total, 38 genes were reported in at least five cancer versus normal studies, of which 22 were up-regulated (Table 3.2, page 50). and the remaining 16 were down-regulated (Table 3.3, page 54). Also, 125 genes were reported in three or four cancer versus normal studies, of which 77 were up-regulated (Table 3.4, page 57) and the remaining 48 genes were down-regulated (Table 3.5, page 64).

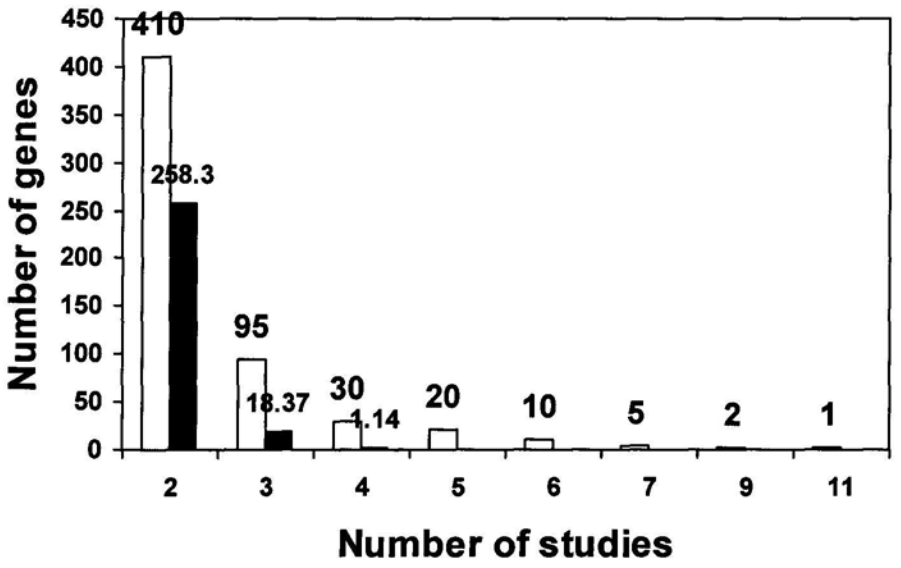

Figure 3.1: Overlap analysis results for the cancer versus normal comparison.

The actual overlap values are represented by white bars, while the average overlap values from the Monte Carlo simulations are represented by black bars. The 95% confidence intervals of the simulations were not included as they were too small to visualize. Note that the average overlap values in the simulations never equaled or exceeded that which was observed in the actual data.

3.3 Overlap Results for Adenoma versus Normal Tissue Comparisons

In the adenoma versus normal tissue comparison, there were 39 multi-study genes, of which 23 were up-regulated (Table 3.6, page 68) and the remaining 16 were down-regulated (Table 3.7 on page 70). In the simulations, an average of 10.64 (95% CI, 10.61 to 10.68) genes was observed with an overlap of two, while the actual data contained 37. For an overlap of three, an average of 0.07 (95% CI, 0.067 to 0.073) of a gene was observed in the simulations, while two genes were observed with an overlap of three in the real data.

3.4 Overlap Results for Cancer versus Adenoma Tissue Comparisons

Finally, in the cancer versus adenoma tissue comparison, there were five multi-study genes; all five were reported in only two independent studies. In the 10,000 permutations, there were 775 instances in which five genes were observed with an overlap of two or more, which yields a non-significant P-value of 0.08 (results not shown).

3.5 Overlapping Results in the Cancer versus Normal and Adenoma versus Normal Comparisons

The cancer versus normal and adenoma versus normal comparisons produced differentially expressed genes reported at a statistically significant frequency. Thus, we determined the overlap between these two comparisons. The 26 genes are presented in Table 3.8, page 72.

Table 3.1: Summary of Comparisons Studied for Overlap Significance

Note the overlap observed in the cancer versus adenoma comparison was not significant (P < .05).

Comparison	Total Number of Studies	Total Number of Differentially Expressed Genes Reported (Mapped)	Total Number of Differentially Expressed Genes Reported in ≥ 2 Studies	P-value
Cancer versus Normal	23	6537 (5886)	573	< .0001
Adenoma versus Normal	7	1101 (986)	39	< .0001
Cancer versus Adenoma	5	538 (415)	5	.08

Table 3.2: Up-regulated Genes Most Commonly Reported in Cancer versus Normal Expression Profiling Studies

The 22 up-regulated genes reported in at least five independent studies with consistent direction are presented here. Genes reported by five and six studies were further sub-divided into semi-quantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold-change criteria for ranking genes when total sample numbers were similar. Validation studies that report a gene as differentially expressed in the opposite direction from that of the meta-analysis are marked with an '*' Abbreviations: FC, fold change; RT-PCR, reverse transcriptase polymerase chain reaction; IHC, immunohistochemistry; NB, northern blot; WB, western blot.

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
TGFβI	Transforming growth factor, beta-induced, 68kDa	9 [1-8] (8)	369 (329)	8.94	1.11 to 32.00	RT-PCR [1, 2, 6, 8, 9]
IFITM1	Interferon induced transmembrane protein 1 (9-27)	9 [2, 3, 6, 8-13] (4)	351 (187)	7.52	3.00 to 12.00	RT-PCR [8, 9]
MYC	V-myc myelocytoma-tosis viral oncogene homolog (avian)	7 [3, 6, 7, 9, 13-15] (4)	329 (243)	5.02	1.69 to 7.50	RT-PCR [6, 9, 13, 16]
SPARC	Secreted protein, acidic, cysteine-rich (osteonectin)	7 [2, 3, 5, 7, 9, 13, 15] (5)	244 (180)	6.30	1.27 to 15.00	IHC [17]*
GDF15	Growth differentiation factor 15	7 [1, 2, 5, 7, 13, 18] (5)	230 (172)	7.42	1.58 to 12.20	RT-PCR [2, 9]

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
6 studies: Greatest sample size						
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	6 [3, 6, 7, 13, 18, 19] (4)	287 (229)	6.54	2.74 to 10.50	RT-PCR [7, 18]
6 studies: Moderate sample size						
CDC25B	Cell division cycle 25 homolog B (S. pombe)	6 [3, 7-9, 13, 19] (4)	256 (176)	4.93	1.81 to 9.20	RT-PCR [19]
HMBG1	High-mobility group box 1	6 [1, 5, 6, 10, 12, 13] (3)	264 (161)	3.27	2.66 to 3.91	WB, IHC [20]
6 studies: Lowest sample size						
IFITM2	Interferon induced transmem-brane protein 2 (1-8D)	6 [1-3, 8, 11, 15] (3)	141 (56)	7.09	3.00 to 13.00	RT-PCR [21]
COL1A2	Collagen, type I, alpha 2	6 [2, 3, 5, 7, 12, 15] (4)	172 (130)	6.93	2.96 to 12.00	None found

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
5 studies: Greatest sample size						
CKS2	CDC28 protein kinase regulatory subunit 2	5 [5-7, 9, 19] (5)	285 (285)	4.21	1.79 to 7.20	RT-PCR [9, 19]
TOP2A	Topoisome-rase (DNA) II alpha 170kDa	5 [6, 7, 9, 13, 22] (4)	277 (237)	3.61	1.05 to 5.60	NB, WB [23]
UBE2C	Ubiquitin-conjugating enzyme E2C	5 [3, 5-7, 10] (4)	274 (229)	3.03	1.48 to 5.00	RT-PCR [24]
5 studies: Moderate sample size						
CDH3	Cadherin 3, type 1, P-cadherin (placental)	5 [1, 3, 7, 9, 25] (5)	194 (194)	18.16	2.78 to 74.00	WB [26, 27]
INHBA	Inhibin, beta A (activin A, activin AB alpha polypeptide)	5 [3, 5, 7, 13, 25] (4)	198 (158)	11.05	1.71 to 37.00	RT-PCR [27]
SLC12A2	Solute carrier family 12 (sodium/ potassium/ chloride transporters), member 2	5 [2, 6, 10, 15, 25] (3)	208 (139)	10.58	3.58 to 15.15	RT-PCR [6]
MMP11	Matrix metallopep-tidase 11 (stromelysin 3)	5 [3, 5, 7, 9, 25] (5)	208 (208)	4.22	1.74 to 5.70	WB, IHC [28]

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
CSE1L	CSE1 chromosome segregation 1-like (yeast)	5 [3, 5, 7, 10, 19] (4)	207 (162)	3.74	1.14 to 5.00	None found
HNRPA1	Heterogeneous nuclear ribonucleo-protein A1	5 [2, 5-8] (4)	243 (203)	2.89	1.01 to 4.50	RT-PCR [29]
5 studies: Lowest sample size						
CDK10	Cyclin-dependent kinase (CDC2-like) 10	5 [2, 3, 7, 22, 25] (5)	150 (150)	13.85	2.66 to 17.59	None found
COL3A1	Collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	5 [4, 5, 7, 12, 13] (3)	178 (120)	4.31	1.24 to 9.38	RT-PCR [12]
COL4A1	Collagen, type IV, alpha 1	5 [3, 5, 7, 12, 15] (3)	168 (126)	2.70	1.05 to 4.00	None found

Table 3.3: Down-regulated Genes Most Commonly Reported in Cancer versus Normal Expression Profiling Studies

The 16 down-regulated genes reported in at least five independent studies with consistent direction are presented here. Genes reported by five and six studies were further sub-divided into semi-quantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold-change criteria for ranking genes when total sample numbers were similar. Validation studies that report a gene as differentially expressed in the opposite direction from that of the meta-analysis are marked with an '*' Abbreviations: FC, fold change; RT-PCR, reverse transcriptase polymerase chain reaction; IHC, immunohistochemistry; NB, northern blot; WB, western blot; MS, mass spectrometry.

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
CA2	Carbonic anhydrase II	11 [1-3, 5, 7, 8, 12, 13, 15, 19, 30] (7)	474 (352)	-15.51	-56.00 to -2.30	RT-PCR [12, 13]
MALL	Mal, T-cell differentiation protein-like	7 [2, 3, 7-9, 15, 19] (5)	244 (180)	-5.34	-10.50 to -1.70	None found
CEACAM1	Carcinoembr-yonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	7 [1, 2, 5, 7, 13, 15, 19] (5)	222 (158)	-10.40	-40.00 to -1.38	RT-PCR [13, 19]
6 studies: Greatest sample size						
HSD11B2	Hydroxysteroid (11-β) dehydrogenase 2	6 [1, 3, 7, 8, 19] (5)	224 (184)	-4.47	-7.60 to -2.23	NB [31]
6 studies: Moderate sample size						
SLC26A2	Solute carrier family 26 (sulfate transporter), member 2	6 [1, 3, 5, 7, 15, 18] (4)	190 (148)	-6.78	-9.09 to -4.04	None found

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
FCGBP	Fc fragment of IgG binding protein	6 [2, 3, 5, 7, 8, 10] (4)	215 (130)	-4.88	-7.00 to -1.31	None found
6 studies: Lowest sample size						
ACADS	Acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	6 [1-3, 5, 13, 19] (5)	168 (128)	-7.11	-20.00 to -2.00	None found
CKB	Creatine kinase, brain	6 [2, 3, 5, 7, 8, 12] (4)	188 (130)	-3.11	-5.00 to -1.10	WB [32]
5 studies: Greatest sample size						
CLU	Clusterin	5 [4, 7, 12, 13, 19] (3)	178 (120)	-3.83	-5.60 to -1.10	IHC [33]
CES2	Carboxyleste-rase 2 (intestine, liver)	5 [3, 5, 7, 15, 19] (4)	186 (162)	-3.58	-6.30 to -1.15	None found
5 studies: Moderate sample size						
CA1	Carbonic anhydrase I	5 [2, 3, 5, 8, 19] (4)	146 (106)	-36.90	-59.00 to -5.30	RT-PCR [8]
GPA33	Glycoprotein A33 (transmem-brane)	5 [1, 2, 7, 11, 15] (5)	131 (86)	-12.51	-32.50 to -1.70	None found
KRT20	Keratin 20	5 [2, 5, 7, 8, 19] (4)	176 (136)	-8.31	-20.40 to -1.65	None found
SELENBP1	Selenium binding protein 1	5 [2, 3, 5, 7, 15] (4)	154 (130)	-2.80	-3.45 to -1.11	WB, IHC, MS [34]

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range	Validation
5 studies: Lowest sample size						
CA12	Carbonic anhydrase XII	5 [1, 2, 5, 8, 15] (3)	126 (62)	-4.41	-7.69 to -2.50	IHC [35]*
FABP1	Fatty acid binding protein 1, liver	5 [2, 3, 8, 12, 15] (2)	116 (34)	-4.28	-5.56 to -3.00	RT-PCR [8]

Table 3.4: Up-regulated Genes Reported in Three or Four Cancer versus Normal Expression Profiling Studies

The 77 up-regulated genes reported by three or four studies were further sub-divided into semi-quantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold-change criteria for ranking genes when total sample numbers were similar. Abbreviation: FC, fold change.

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
4 Studies: Greatest sample size					
PCNA	Proliferating cell nuclear antigen	4 [5-7, 9] (4)	249 (249)	3.34	1.20 to 4.80
GTF3A	General transcription factor IIIA	4 [5-7, 19] (4)	235 (235)	3.03	1.50 to 3.35
SOX4	SRY (sex determining region Y)-box 4	4 [5-7, 10] (3)	244 (199)	2.41	2.02 to 2.72
SOX9	SRY (sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	4 [5-7, 10] (3)	280 (235)	2.36	1.89 to 2.25
VEGFA	Vascular endothelial growth factor A	4 [7, 14, 15, 30] (2)	270 (224)	2.07	1.81 to 2.33
4 Studies: Moderate sample size					
ETV4	Ets variant gene 4 (E1A enhancer binding protein, E1AF)	4 [5, 6, 9, 10] (3)	234 (189)	32.64	2.08 to 92.00
MMP1	Matrix metallopepti-dase 1 (interstitial collagenase)	4 [3, 5, 7, 19] (4)	162 (162)	8.26	3.47 to 21.00
CXCL3	Chemokine (C-X-C motif) ligand 3	4 [3, 5, 7, 19] (4)	162 (162)	7.28	2.48 to 13.00

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
NPM1	Nucleophos-min (nucleolar phosphor-protein B23, numatrin)	4 [1, 2, 6, 10] (3)	174 (129)	7.10	4.27 to 12.00
HMGA1	High mobility group AT-hook 1	4 [5, 6, 9, 13] (3)	229 (189)	5.50	2.90 to 7.09
RPS2	Ribosomal protein S2	4 [1, 2, 6, 8] (3)	169 (129)	5.16	2.39 to 8.13
RPL8	Ribosomal protein L8	4[2, 5, 6, 8] (3)	183 (143)	3.50	2.16 to 5.35
EIF3S9	Eukaryotic translation initiation factor 3, subunit 9 eta, 116kDa	4 [5, 6, 8, 9] (3)	229 (189)	3.29	2.53 to 4.30
PRKDC	Protein kinase, DNA-activated, catalytic polypeptide	4 [5-7, 14] (3)	221 (199)	2.49	1.33 to 3.23
4 Studies: Lowest sample size					
MMP7	Matrix metallopepti-dase 7 (matrilysin, uterine)	4 [3, 5, 7, 25] (4)	158 (158)	9.39	3.32 to 21.00
HSPD1	Heat shock 60kDa protein 1 (chaperonin)	4 [1, 4, 15, 36] (3)	82 (58)	4.90	4.12 to 5.99
RAN	RAN, member RAS oncogene family	4 [1, 8-10] (2)	157 (72)	4.60	3.30 to 5.90
PABPC1	Poly(A) binding protein, cytoplasmic 1	4 [1, 2, 5, 12] (3)	80 (62)	3.99	3.00 to 5.75

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
HSP90AB1	Heat shock protein 90kDa alpha (cytosolic), class B member 1	4 [1, 4, 5, 15] (3)	106 (82)	3.72	2.87 to 5.34
3 Studies: Greatest sample size					
RPL29	Ribosomal protein L29	3 [6, 8, 10] (1)	188 (103)	5.35	5.35 to 5.35
RRM2	Ribonucleotide reductase M2 polypeptide	3 [5, 6, 9] (3)	189 (189)	4.85	2.06 to 6.80
RPS7	Ribosomal protein S7	3 [5, 6, 10] (2)	184 (139)	3.96	2.21 to 5.71
MIF	Macrophage migration inhibitory factor (glycosylation-inhibiting factor)	3 [5, 6, 8] (2)	179 (139)	3.68	2.62 to 4.74
FPRL1	Formyl peptide receptor-like 1	3 [6, 7, 14] (3)	195 (195)	3.67	2.92 to 5.15
ENC1	Ectodermal-neural cortex (with BTB-like domain)	3 [5, 25, 30] (3)	232 (232)	3.34	2.62 to 3.92
BMP4	Bone morphogenetic protein 4	3 [5, 6, 13] (2)	179 (139)	3.24	2.92 to 3.55
CPNE1	Copine I	3 [4, 6, 7] (3)	187 (187)	3.04	1.48 to 4.41
POLR1D	Polymerase (RNA) I polypeptide D, 16kDa	3 [5, 6, 13] (2)	179 (139)	2.51	2.10 to 2.91
TGIF1	TGFB-induced factor (TALE family homeobox)	3 [5-7] (3)	199 (199)	2.40	1.38 to 3.60
TRAP1	TNF receptor-associated protein 1	3 [5, 6, 13] (2)	179 (139)	2.28	2.17 to 2.38
MCM3	MCM3 minichromosome maintenance deficient 3 (S. cerevisiae)	3 [5-7] (3)	199 (199)	2.22	1.30 to 2.98

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
SNRPB	Small nuclear ribonucleo-protein polypeptides B and B1	3 [5-7] (3)	199 (199)	2.18	1.01 to 3.03
ITGA2	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	3 [10, 14, 30] (1)	231 (164)	2.06	2.06 to 2.06
TOMM40	Translocase of outer mitochondrial membrane 40 homolog (yeast)	3 [5-7] (3)	199 (199)	1.94	1.66 to 2.13
3 Studies: Moderate sample size					
FN1	Fibronectin 1	3 [2, 7, 9] (3)	114 (114)	9.94	1.11 to 15.00
SPP1	Secreted phosphor-protein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	3 [3, 5, 7] (3)	126 (126)	6.23	2.23 to 12.00
TACSTD2	Tumor-associated calcium signal transducer 2	3 [5, 7, 25] (3)	128 (128)	6.04	4.53 to 7.84
BGN	Biglycan	3 [3, 6, 13] (2)	173 (133)	5.23	4.46 to 6.00
WEE1	WEE1 homolog (S. pombe)	3 [4, 7, 9] (3)	134 (134)	4.70	1.16 to 6.75
RPS19	Ribosomal protein S19	3 [2, 6, 8] (2)	147 (107)	4.59	4.05 to 5.13
SERPINE1	Serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	3 [3, 7, 14] (2)	112 (90)	4.32	1.64 to 7.00

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
CXCL2	Chemokine (C-X-C motif) ligand 2	3 [3, 5, 7] (3)	126 (126)	4.23	2.21 to 8.00
FCGR3A	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)	3 [7, 9, 11] (2)	131 (110)	4.22	2.73 to 5.70
RPL31	Ribosomal protein L31	3 [6, 10, 12] (1)	166 (103)	4.08	4.08 to 4.08
RPS5	Ribosomal protein S5	3 [2, 6, 8] (2)	147 (107)	3.79	2.67 to 4.90
HSPE1	Heat shock 10kDa protein 1 (chaperonin 10)	3 [5, 6, 19] (3)	175 (175)	3.74	2.54 to 4.80
EIF3S6	Eukaryotic translation initiation factor 3, subunit 6 48kDa	3 [1, 6, 10] (2)	170 (125)	3.73	3.26 to 4.20
HOMER1	Homer homolog 1 (Drosophila)	3 [3, 6, 25] (3)	165 (165)	3.65	2.55 to 5.00
ODC1	Ornithine decarboxylase 1	3 [1, 6, 8] (2)	165 (125)	3.59	3.27 to 3.90
HMGB2	High-mobility group box 2	3 [1, 6, 10] (2)	170 (125)	3.42	2.30 to 4.54
PYCR1	Pyrroline-5-carboxylate reductase 1	3 [7, 8, 19] (2)	136 (96)	3.27	1.23 to 5.30
RPS18	Ribosomal protein S18	3 [2, 4, 6] (3)	131 (131)	3.25	2.02 to 4.80
TRIM28	Tripartite motif-containing 28	3 [2, 5, 6] (3)	143 (143)	2.90	2.02 to 4.57
SORD	Sorbitol de-hydrogenase	3 [5, 7, 19] (3)	132 (132)	2.86	1.19 to 4.70
HNRPH1	Heterogeneous nuclear ribonucleoprotein H1 (H)	3 [6, 11, 18] (1)	142 (103)	2.81	2.81 to 2.81
PPIB	Peptidylprolyl isomerase B (cyclophilin B)	3 [5, 6, 37] (2)	149 (139)	2.63	2.50 to 2.75
IMPDH2	IMP (inosine monophosphate) dehydrogenase 2	3 [1, 6, 8] (2)	165 (125)	2.57	2.32 to 2.81

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
GSTP1	Glutathione S-transferase pi	3 [5, 6, 12] (2)	157 (139)	2.43	2.38 to 2.48
C2	Complement component 2	3 [3, 7, 8] (2)	130 (90)	2.21	1.42 to 3.00
GGH	Gamma-glutamyl hydrolase (conjugase, folylpoly-gammaglut-amylhydrolase)	3 [5, 7, 13] (2)	136 (96)	1.87	1.47 to 2.27
CDKN3	Cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)	3 [5, 7, 11] (2)	117 (96)	1.80	1.11 to 2.48
MYBL2	V-myb myeloblastosis viral oncogene homolog (avian)-like 2	3 [7, 11, 13] (1)	121 (60)	1.69	1.69 to 1.69
3 Studies: Lowest sample size					
GNB2L1	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	3 [2, 10, 37] (1)	59 (4)	8.30	8.30 to 8.30
COL1A1	Collagen, type I, alpha 1	3 [2, 7, 15] (2)	88 (64)	8.20	2.39 to 14.00
CLDN2	Claudin 2	3 [2, 5, 15] (2)	64 (40)	7.44	3.88 to 11.00
BST2	Bone marrow stromal cell antigen 2	3 [2, 5, 7] (3)	100 (100)	6.82	1.19 to 3.28
AHCY	S-adenosyl-homocysteine hydrolase	3 [2, 3, 19] (3)	70 (70)	5.78	4.00 to 7.33
LDHB	Lactate dehydro-genase B	3 [1, 5, 8] (2)	98 (58)	4.26	2.05 to 6.47
RPL30	Ribosomal protein L30	3 [1, 8, 12] (1)	80 (22)	4.10	4.10 to 4.10
CFB	Complement factor B	3 [13, 16, 18] (1)	94 (36)	3.70	3.70 to 3.70

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
RPL6	Ribosomal protein L6	3 [1, 8, 10] (1)	107 (22)	3.46	3.46 to 3.46
BMP7	Bone morphogenetic protein 7 (osteogenic protein 1)	3 [11, 13, 25] (1)	93 (32)	3.42	3.42 to 3.42
NME1	Non-metastatic cells 1, protein (NM23A) expressed in	3 [2, 5, 10] (2)	85 (40)	3.37	2.57 to 4.17
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	3 [1, 10, 37] (1)	77 (22)	3.12	3.12 to 3.12
CCT7	Chaperonin containing TCP1, subunit 7 (eta)	3 [1, 5, 8] (2)	98 (58)	2.94	2.40 to 3.47
RPSA	Ribosomal protein SA	3 [2, 10, 37] (1)	59 (4)	2.74	2.74 to 2.74
RPL3	Ribosomal protein L3	3 [1, 2, 8] (2)	66 (26)	2.33	2.30 to 2.36

Table 3.5: Down-regulated Genes Reported in Three or Four Cancer versus Normal Expression Profiling Studies

The 48 down-regulated genes reported by three or four studies were further sub-divided into semi-quantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold-change criteria for ranking genes when total sample numbers were similar. Abbreviation: FC, fold change.

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
4 Studies: Greatest sample size					
CD177	CD177 molecule	4 [3, 5, 6, 15] (3)	193 (169)	-15.09	-30.00 to -6.95
SEPP1	Selenoprotein P, plasma, 1	4 [4-6, 15] (3)	187 (163)	-4.56	-39.00 to -12.5
CNN1	Calponin 1, basic, smooth muscle	4 [5, 7, 10, 12] (2)	159 (96)	-2.91	-4.00 to -1.82
4 Studies: Moderate sample size					
CA4	Carbonic anhydrase IV	4 [3, 5, 10, 19] (3)	147 (102)	-29.80	-4.00 to -1.21
MT1H	Metallothionein 1H	4 [5, 12, 13, 19] (2)	130 (72)	-6.65	-7.69 to -5.60
ADH1C	Alcohol dehydrogenase 1C (class I), gamma polypeptide	4 [1, 3, 7, 19] (4)	148 (148)	-4.88	-10.00 to -2.08
VIPR1	Vasoactive intestinal peptide receptor 1	4 [1, 5, 8, 19] (3)	134 (94)	-4.50	-8.00 to -2.17
MYH11	Myosin, heavy chain 11, smooth muscle	4 [3, 7, 10, 12] (2)	153 (90)	-3.11	-4.00 to -2.21
4 Studies: Lowest sample size					
ITM2C	Integral membrane protein 2C	4 [1, 2, 5, 15] (3)	86 (62)	-4.39	-6.95 to -30.00
HIGD1A	HIG1 domain family, member 1A	4 [2, 5, 10, 15] (2)	109 (40)	-4.37	-5.50 to -3.23

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
FXYD3	FXYD domain containing ion transport regulator 3	4 [2, 3, 7, 15] (3)	118 (94)	-2.82	-4.00 to -1.21
3 Studies: Greatest sample size					
SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	3 [4, 9, 10] (2)	119 (74)	-18.57	-33.33 to -3.80
MS4A12	Membrane-spanning 4-domains, subfamily A, member 12	3 [2, 5, 6] (3)	143 (143)	-17.11	-36.00 to -7.14
SLC4A4	Solute carrier family 4, sodium bicarbonate cotransporter, member 4	3 [3, 5, 6] (3)	169 (169)	-11.25	-25.00 to -3.03
ANPEP	Alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)	3 [3, 5, 9] (3)	116 (116)	-7.08	-13.00 to -3.23
MT1G	Metallothionein 1G	3 [12, 19, 30] (3)	218 (200)	-3.15	-4.10 to -2.20
UGT1A6	UDP glucuronosyltransferase 1 family, polypeptide A6	3 [5, 7, 8] (2)	136 (96)	-2.90	-3.45 to -2.34
MYL9	Myosin, light chain 9, regulatory	3 [5, 7, 8] (2)	120 (96)	-1.88	-2.50 to -1.26
TST	Thiosulfate sulfurtransferase (rhodanese)	3 [5, 7, 8] (2)	136 (96)	-1.73	-2.27 to -1.19
PRDX6	Peroxiredoxin 6	3 [5, 7, 13] (2)	136 (96)	-1.72	-2.33 to -1.10
3 Studies: Moderate sample size					
MGLL	Monoglyceride lipase	3 [2, 5, 10] (2)	85 (40)	-19.52	-36.00 to -3.03
CHGA	Chromogranin A (parathyroid secretory protein 1)	3 [3, 10, 19] (2)	111 (66)	-12.25	-19.00 to -5.50

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
GSN	Gelsolin (amyloidosis, Finnish type)	3 [2, 5, 10] (2)	85 (40)	-9.96	-17.54 to -2.38
TSPAN1	Tetraspanin 1	3 [2, 5, 8] (2)	80 (40)	-9.75	-14.50 to -5.00
HSD17B2	Hydroxysteroid (17-beta) dehydrogenase 2	3 [3, 5, 19] (3)	102 (102)	-9.59	-18.00 to -5.56
EDN3	Endothelin 3	3 [3, 5, 13] (2)	106 (66)	-7.50	-10.00 to -5.00
SMPDL3A	Sphingomyelin phosphodiesterase, acid-like 3A	3 [2, 5, 7] (3)	100 (100)	-6.55	-15.00 to -1.41
PLS1	Plastin 1 (I isoform)	3 [2, 5, 7] (3)	100 (100)	-6.32	-15.50 to -1.12
MT1A	Metallothionein 1A (functional)	3 [12, 13, 19] (1)	94 (36)	-5.70	-5.70 to -5.70
MEP1A	Meprin A, alpha (PABA peptide hydrolase)	3 [2, 5, 19] (3)	76 (76)	-5.56	-7.10 to -4.83
ABP1	Amiloride binding protein 1 (amine oxidase (copper-containing))	3 [5, 8, 19] (2)	112 (72)	-5.32	-8.20 to -2.44
APBA3	Amyloid beta (A4) precursor protein-binding, family A, member 3 (X11-like 2)	3 [1, 2, 10] (2)	71 (26)	-3.92	-4.50 to -3.33
TSPAN7	Tetraspanin 7	3 [1, 5, 15] (2)	82 (58)	-3.45	-3.57 to -3.33
NCAM2	Neural cell adhesion molecule 2	3 [1, 5, 8] (2)	98 (58)	-2.75	-3.33 to -2.17
HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	3 [5, 7, 12] (2)	114 (96)	-2.65	-3.85 to -1.45
C1orf115	Chromosome 1 open reading frame 115	3 [5, 13, 18] (1)	94 (36)	-2.13	-2.13 to -2.13
CCNYL1	Hypothetical protein FLJ40432	3 [5, 8, 18] (1)	94 (36)	-2.08	-2.08 to -2.08

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	3 [3, 8, 15] (1)	94 (30)	-1.00	-1.00 to -1.00
3 Studies: Lowest sample size					
GUCA2A	Guanylate cyclase activator 2A (guanylin)	3 [2, 3, 19] (3)	70 (70)	-27.68	-52.00 to -10.85
MT2A	Metallothionein 2A	3 [2, 5, 12] (2)	58 (40)	-11.02	-19.00 to -3.03
MUC2	Mucin 2, oligomeric mucus/gel-forming	3 [2, 5, 11] (2)	61 (40)	-9.59	-15.00 to -4.17
CLCA1	Chloride channel, calcium activated, family member 1	3 [2, 3, 5] (3)	70 (70)	-8.56	-11.00 to -7.00
GCNT3	Glucosaminyl (N-acetyl) transferase 3, mucin type	3 [2, 5, 15] (2)	64 (40)	-7.38	-10.00 to -4.76
TMEM54	Transmembrane protein 54	3 [2, 12, 18] (1)	40 (4)	-7.00	-7.00 to -7.00
MUC12	Mucin 12, cell surface associated	3 [2, 5, 15] (2)	64 (40)	-4.79	-6.00 to -3.57
KRT8	Keratin 8	3 [2, 12, 15] (1)	46 (4)	-4.22	-4.22 to -4.22
KRT17	Keratin 17	3 [1, 2, 12] (2)	44 (26)	-3.89	-4.45 to -3.33
LGALS4	Lectin, galactoside-binding, soluble, 4 (galectin 4)	3 [2, 8, 15] (1)	68 (4)	-3.34	-3.34 to -3.34

Table 3.6: Up-regulated Genes Most Commonly Reported in Adenoma versus Normal Expression Profiling Studies.

The 23 up-regulated genes reported by two or three studies were further sub-divided into semi-quantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold-change criteria for ranking genes when total sample numbers were similar. Abbreviation: FC, fold change.

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
RPS3	Ribosomal protein S3	3 [1, 2, 19] (3)	30 (30)	4.345	3.48 to 5.21
2 Studies: Greatest sample size					
EPHB2	EPH receptor B2	2 [2, 16] (2)	37 (37)	7.55	7.10 to 8.00
SLC12A2	Solute carrier family 12 (sodium/potass-ium/chloride transporters), member 2	2 [2, 16] (2)	37 (37)	5.44	5.30 to 5.57
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	2 [1, 37] (2)	24 (24)	2.65	2.18 to 3.12
2 Studies: Moderate sample size					
ETS2	V-ets erythroblastosis virus E26 oncogene homolog 2 (avian)	2 [2, 19] (1)	12 (4)	20.00	20.00 to 20.00
GDF15	Growth differentiation factor 15	2 [1, 2] (2)	22 (22)	12.25	7.50 to 17.00
PABPC1	Poly(A) binding protein, cytoplasmic 1	2 [1, 2] (2)	22 (22)	11.87	3.23 to 20.50
HMGB1	High-mobility group box 1	2 [1, 2] (2)	22 (22)	9.12	3.23 to 15.00
RPS29	Ribosomal protein S29	2 [2, 19] (1)	12 (4)	7.05	7.05 to 7.05
RPS27A	Ribosomal protein S27a	2 [1, 2] (2)	22 (22)	6.69	2.60 to 10.77

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
NPM1	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)	2 [1, 2] (2)	22 (22)	6.27	5.03 to 7.50
RNF43	Ring finger protein 43	2 [1, 2] (2)	22 (22)	6.22	4.43 to 8.00
ZFP36L1	Zinc finger protein 36, C3H type-like 1	2 [1, 2] (2)	22 (22)	5.91	2.81 to 9.00
IMPDH2	IMP (inosine monophosphate) dehydrogenase 2	2 [1, 2] (2)	22 (22)	5.66	2.81 to 8.50
RPL30	Ribosomal protein L30	2 [1, 2] (2)	22 (22)	4.30	4.10 to 4.49
RPS15	Ribosomal protein S15	2 [2, 19] (1)	12 (4)	3.90	3.90 to 3.90
RPS4X	Ribosomal protein S4, X-linked	2 [1, 2] (2)	22 (22)	3.77	2.93 to 4.60
RPS2	Ribosomal protein S2	2 [1, 2] (2)	22 (22)	3.01	2.06 to 3.96
TUBA3	Tubulin, alpha 3	2 [1, 2] (2)	22 (22)	3.00	2.95 to 3.05
RPS25	Ribosomal protein S25	2 [2, 19] (1)	12 (4)	2.81	2.81 to 2.81
RPL3	Ribosomal protein L3	2 [1, 2] (2)	22 (22)	2.20	2.04 to 2.36
2 Studies: Lowest sample size					
GNB2L1	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	2 [2, 37] (2)	10 (10)	4.65	3.90 to 5.40
RPSA	Ribosomal protein SA	2 [2, 37] (2)	10 (10)	3.45	2.31 to 4.58

Table 3.7: Down-regulated Genes Most Commonly Reported in Adenoma versus Normal Expression Profiling Studies

The 16 down-regulated genes reported by two or three studies were further sub-divided into semi-quantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold-change criteria for ranking genes when total sample numbers were similar. Abbreviation: FC, fold change.

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
KRT8	Keratin 8	3 [2, 19, 38] (3)	38 (38)	-3.45	-5.14 to -2.36
2 Studies: Greatest sample size					
SEPP1	Selenoprotein P, plasma, 1	2 [2, 16] (2)	37 (37)	-10.67	-18.00 to -3.33
IL12RB1	Interleukin 12 receptor, beta 1	2 [1, 22] (2)	40 (40)	-2.70	-3.33 to -2.06
CETP	Cholesteryl ester transfer protein, plasma	2 [1, 16] (2)	51 (51)	-2.67	-3.33 to -2.00
ITPKB	Inositol 1,4,5-trisphosphate 3-kinase B	2 [1, 16] (2)	51 (51)	-2.50	-3.33 to -1.67
2 Studies: Lowest/ Moderate sample size					
CA2	Carbonic anhydrase II	2 [1, 2] (2)	22 (22)	-16.50	-28.00 to -5.00
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	2 [1, 2] (2)	22 (22)	-15.00	-27.50 to -2.50
GPA33	Glycoprotein A33 (transmembrane)	2 [1, 2] (2)	22 (22)	-12.50	-21.67 to -3.33
SLC26A3	Solute carrier family 26, member 3	2 [1, 2] (2)	22 (22)	-12.17	-21.00 to -3.33
FKBP1A	FK506 binding protein 1A, 12kDa	2 [1, 2] (2)	22 (22)	-10.00	-15.00 to -5.00
KRT17	Keratin 17	2 [1, 2] (2)	22 (22)	-6.58	-9.83 to -3.33
THBS2	Thrombospondin 2	2 [1, 38] (2)	30 (30)	-5.70	-8.90 to -2.50
PRSS8	Protease, serine, 8 (prostasin)	2 [1, 2] (2)	22 (22)	-5.65	-8.80 to -2.50
KIAA0828	Adenosylhomocysteinase 3	2 [1, 2] (2)	22 (22)	-4.54	-5.75 to -3.33

Gene Name	Description	Studies (with FC)	Total Sample Sizes (with FC)	Mean Fold Change	Range
ACADS	Acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	2 [1, 2] (2)	22 (22)	-4.33	-5.33 to -3.33
UQCR	Ubiquinol-cytochrome c reductase, 6.4kDa subunit	2 [1, 2] (2)	22 (22)	-3.36	-4.22 to -2.50

Table 3.8: Differentially Expressed Genes Reported with a Statistically Significant Frequency in the Adenoma versus Normal and Cancer versus Normal Comparisons

Gene Name	Description	Direction of Differential Expression	Overlap in Adenoma versus Normal	Overlap in Cancer versus Normal
GDF15	Growth differentiation factor 15	UP	2	7
HMGB1	High-mobility group box 1	UP	2	6
SLC12A2	Solute carrier family 12 (sodium/potassium/chloride transporters), member 2	UP	2	5
NPM1	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)	UP	2	4
IMPDH2	IMP (inosine monophosphate) dehydrogenase 2	UP	2	4
PABPC1	Poly(A) binding protein, cytoplasmic 1	UP	2	4
RNF43	Ring finger protein 43	UP	2	3
GNB2L1	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	UP	2	3
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	UP	2	3
RPL3	Ribosomal protein L3	UP	2	3
RPSA	Ribosomal protein SA	UP	2	3
RPL30	Ribosomal protein L30	UP	2	3
RPS3	Ribosomal protein S3	UP	3	2
RPS2	Ribosomal protein S2	UP	2	2
RPS4X	Ribosomal protein S4, X-linked	UP	2	2
RPS15	Ribosomal protein S15	UP	2	2
CA2	Carbonic anhydrase II	DOWN	2	11
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	DOWN	2	7
ACADS	Acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	DOWN	2	6
GPA33	Glycoprotein A33 (transmembrane)	DOWN	2	5
SEPP1	Selenoprotein P, plasma, 1	DOWN	2	4
KRT8	Keratin 8	DOWN	3	3
KRT17	Keratin 17	DOWN	2	3
ITPKB	Inositol 1,4,5-trisphosphate 3-kinase B	DOWN	2	2
PRSS8	Protease, serine, 8 (prostasin)	DOWN	2	2
UQCR	Ubiquinol-cytochrome c reductase, 6.4kDa subunit	DOWN	2	2

Chapter 3 References

1. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y: **Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas.** *Oncogene* 2002, **21:**4120-4128.

2. Buckhaults P, Rago C, St Croix B, Romans KE, Saha S, Zhang L, Vogelstein B, Kinzler KW: **Secreted and cell surface genes expressed in benign and malignant colorectal tumors.** *Cancer Res* 2001, **61:**6996-7001.

3. Croner RS, Foertsch T, Brueckl WM, Guenther K, Siebenhaar R, Stremmel C, Matzel KE, Papadopoulos T, Kirchner T, Behrens J, Klein-Hitpass L, Stuerzl M, Hohenberger W, Reingruber B: **Common denominator genes that distinguish colorectal carcinoma from normal mucosa.** *Int J Colorectal Dis* 2005, **20:**353-362.

4. Kwon HC, Kim SH, Roh MS, Kim JS, Lee HS, Choi HJ, Jeong JS, Kim HJ, Hwang TH: **Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer.** *Dis Colon Rectum* 2004, **47:**141-152.

5. Kim H, Nam SW, Rhee H, Shan Li L, Ju Kang H, Hye Koh K, Kyu Kim N, Song J, Tak-Bun Liu E, Kim H: **Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas.** *Oncogene* 2004, **23:**6218-6225.

6. Grade M, Hormann P, Becker S, Hummon AB, Wangsa D, Varma S, Simon R, Liersch T, Becker H, Difilippantonio MJ, Ghadimi BM, Ried T: **Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas.** *Cancer Res* 2007, **67:**41-56.

7. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ: **Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling.** *J Natl Cancer Inst* 2002, **94:**513-521.

8. Kitahara O, Furukawa Y, Tanaka T, Kihara C, Ono K, Yanagawa R, Nita ME, Takagi T, Nakamura Y, Tsunoda T: **Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia.** *Cancer Res* 2001, **61:**3544-3549.

9. Koehler A, Bataille F, Schmid C, Ruemmele P, Waldeck A, Blaszyk H, Hartmann A, Hofstaedter F, Dietmaier W: **Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage.** *J Pathol* 2004, **204:**65-74.

10. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loriod B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D, Houlgatte R: **Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters.** *Oncogene* 2004, **23:**1377-1391.

11. Ichikawa Y, Ishikawa T, Takahashi S, Hamaguchi Y, Morita T, Nishizuka I, Yamaguchi S, Endo I, Ike H, Togo S, Oki S, Shimada H, Kadota K, Nakamura S, Goto H, Nitanda H, Satomi S, Sakai T, Narita I, Gejyo F, Tomaru Y, Shimizu K, Hayashizaki Y, Okazaki Y: **Identification of genes regulating colorectal carcinogenesis by using the algorithm for diagnosing malignant state method.** *Biochem Biophys Res Commun* 2002, **296:**497-506.

12. Jansova E, Koutna I, Krontorad P, Svoboda Z, Krivankova S, Zaloudik J, Kozubek M, Kozubek S: **Comparative transcriptome maps: A new approach to the diagnosis of colorectal carcinoma patients using cDNA microarrays.** *Clin Genet* 2006, **69:**218-227.

13. Williams NS, Gaynor RB, Scoggin S, Verma U, Gokaslan T, Simmang C, Fleming J, Tavana D, Frenkel E, Becerra C: **Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference.** *Clin Cancer Res* 2003, **9:**931-946.

14. Sugiyama Y, Farrow B, Murillo C, Li J, Watanabe H, Sugiyama K, Evers BM: **Analysis of differential gene expression patterns in colon cancer and cancer stroma using microdissected tissues.** *Gastroenterology* 2005, **128:**480-486.

15. Takemasa I, Higuchi H, Yamamoto H, Sekimoto M, Tomita N, Nakamori S, Matoba R, Monden M, Matsubara K: **Construction of preferential cDNA microarray specialized for human colorectal carcinoma: Molecular sketch of colorectal cancer.** *Biochem Biophys Res Commun* 2001, **285:**1244-1249.

16. Habermann JK, Paulsen U, Roblick UJ, Upender MB, McShane LM, Korn EL, Wangsa D, Kruger S, Duchrow M, Bruch HP, Auer G, Ried T: **Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis.** *Genes Chromosomes Cancer* 2007, **46:**10-26.

17. Yang E, Kang HJ, Koh KH, Rhee H, Kim NK, Kim H: **Frequent inactivation of SPARC by promoter hypermethylation in colon cancers.** *Int J Cancer* 2007, **121:**567-575.

18. Zou TT, Selaru FM, Xu Y, Shustova V, Yin J, Mori Y, Shibata D, Sato F, Wang S, Olaru A, Deacu E, Liu TC, Abraham JM, Meltzer SJ: **Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon.** *Oncogene* 2002, **21:**4855-4862.

19. Notterman DA, Alon U, Sierk AJ, Levine AJ: **Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays.** *Cancer Res* 2001, **61:**3124-3130.

20. Volp K, Brezniceanu ML, Bosser S, Brabletz T, Kirchner T, Gottel D, Joos S, Zornig M: **Increased expression of high mobility group box 1 (HMGB1) is associated with an elevated level of the antiapoptotic c-IAP2 protein in human colon carcinomas.** *Gut* 2006, **55:**234-242.

21. Andreu P, Colnot S, Godard C, Laurent-Puig P, Lamarque D, Kahn A, Perret C, Romagnolo B: **Identification of the IFITM family as a new molecular marker in human colorectal tumors.** *Cancer Res* 2006, **66:**1949-1955.

22. Galamb O, Sipos F, Dinya E, Spisak S, Tulassay Z, Molnar B: **mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray.** *World J Gastroenterol* 2006, **12:**6998-7006.

23. Shibao K, Takano H, Nakayama Y, Okazaki K, Nagata N, Izumi H, Uchiumi T, Kuwano M, Kohno K, Itoh H: **Enhanced coexpression of YB-1 and DNA topoisomerase II alpha genes in human colorectal carcinomas.** *Int J Cancer* 1999, **83:**732-737.

24. Takahashi Y, Ishii Y, Nishida Y, Ikarashi M, Nagata T, Nakamura T, Yamamori S, Asai S: **Detection of aberrations of ubiquitin-conjugating enzyme E2C gene (UBE2C) in advanced colon cancer with liver metastases by DNA microarray and two-color FISH.** *Cancer Genet Cytogenet* 2006, **168:**30-35.

25. Ohmachi T, Tanaka F, Mimori K, Inoue H, Yanaga K, Mori M: **Clinical significance of TROP2 expression in colorectal cancer.** *Clin Cancer Res* 2006, **12:**3057-3063.

26. Hardy RG, Tselepis C, Hoyland J, Wallis Y, Pretlow TP, Talbot I, Sanders DS, Matthews G, Morton D, Jankowski JA: **Aberrant P-cadherin expression is an early event in hyperplastic and dysplastic transformation in the colon.** *Gut* 2002, **50:**513-519.

27. Wildi S, Kleeff J, Maruyama H, Maurer CA, Buchler MW, Korc M: **Overexpression of activin A in stage IV colorectal cancer.** *Gut* 2001, **49:**409-417.

28. Madoz-Gurpide J, Lopez-Serra P, Martinez-Torrecuadrada JL, Sanchez L, Lombardia L, Casal JI: **Proteomics-based validation of genomic data: Applications in colorectal cancer diagnosis.** *Mol Cell Proteomics* 2006, **5:**1471-1483.

29. Ushigome M, Ubagai T, Fukuda H, Tsuchiya N, Sugimura T, Takatsuka J, Nakagama H: **Up-regulation of hnRNP A1 gene in sporadic human colorectal cancers.** *Int J Oncol* 2005, **26:**635-640.

30. Bianchini M, Levy E, Zucchini C, Pinski V, Macagno C, De Sanctis P, Valvassori L, Carinci P, Mordoh J: **Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa.** *Int J Oncol* 2006, **29:**83-94.

31. Takahashi K, Sasano H, Fukushima K, Hirasawa G, Miura H, Sasaki I, Matsuno S, Krozowski ZS, Nagura H: **11 beta-hydroxysteroid dehydrogenase type II in human colon: A new marker of fetal development and differentiation in neoplasms.** *Anticancer Res* 1998, **18:**3381-3388.

32. Balasubramani M, Day BW, Schoen RE, Getzenberg RH: **Altered expression and localization of creatine kinase B, heterogeneous nuclear ribonucleoprotein F, and high mobility group box 1 protein in the nuclear matrix associated with colon cancer.** *Cancer Res* 2006, **66:**763-769.

33. Chen X, Halberg RB, Ehrhardt WM, Torrealba J, Dove WF: **Clusterin as a biomarker in murine and human intestinal neoplasia.** *Proc Natl Acad Sci U S A* 2003, **100:**9530-9535.

34. Kim H, Kang HJ, You KT, Kim SH, Lee KY, Kim TI, Kim C, Song SY, Kim HJ, Lee C, Kim H: **Suppression of human selenium-binding protein 1 is a late event in colorectal carcinogenesis and is associated with poor survival.** *Proteomics* 2006, **6:**3466-3476.

35. Kivela AJ, Saarnio J, Karttunen TJ, Kivela J, Parkkila AK, Pastorekova S, Pastorek J, Waheed A, Sly WS, Parkkila TS, Rajaniemi H: **Differential expression of cytoplasmic carbonic anhydrases, CA I and II, and membrane-associated isozymes, CA IX and XII, in normal mucosa of large intestine and in colorectal tumors.** *Dig Dis Sci* 2001, **46:**2179-2186.

36. Mori D, Nakafusa Y, Miyazaki K, Tokunaga O: **Differential expression of janus kinase 3 (JAK3), matrix metalloproteinase 13 (MMP13), heat shock protein 60 (HSP60), and mouse double minute 2 (MDM2) in human colorectal cancer progression using human cancer cDNA microarrays.** *Pathol Res Pract* 2005, **201:**777-789.

37. Wang JY, Yeh CS, Tzou WS, Hsieh JS, Chen FM, Lu CY, Yu FJ, Cheng TL, Huang TJ, Lin SR: **Analysis of progressively overexpressed genes in tumorigenesis of colorectal cancers using cDNA microarray.** *Oncol Rep* 2005, **14:**65-72.

38. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F: **Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia.** *Gut* 2003, **52:**1148-1153.

Chapter 4: Discussion

4.1 Meta-analyses to Elucidate Biologically Relevant Genes in Colorectal Cancer Tumorigenesis

A logical solution to the problem of lack of agreement between expression profiling studies in colorectal cancer is to determine the overlap between many studies utilizing different platforms and observe which genes are consistently reported as differentially expressed. These genes likely demonstrate biological relevance to the tumorigenesis of colorectal cancer, as opposed to sporadically reported genes, which may be false positives.

Meta-analyses have been previously performed to determine differentially expressed genes in colorectal cancer [1-3]. However, these studies and others usually do not consider whether the level of overlap observed is statistically significant. In the most recent version (3.0) of the cancer profiling database Oncomine [4], a meta-analysis tool was implemented to compare results from independent studies. However, Oncomine presently contains raw data for eight colorectal cancer profiling studies, only two of which would qualify for our study [5, 6] because they were the only studies that performed at least one of the three comparisons of interest. As previously discussed, our meta-analysis method is useful when raw data is unavailable for consistent re-analysis, which is usually the case. Furthermore, in the original meta-analysis study, we were able to obtain raw microarray data from five of the studies included in the meta-analysis

[7]. Consistent re-analysis of these raw data showed that there was a highly significant level agreement between the two methods. However, one limitation of our meta-analysis method is that a measure of confidence cannot be assigned at the gene level, such as from calculating a true combined fold-change or P-value. Thus, in order for more powerful meta-analysis methods to be applied to colorectal cancer profiling studies, researchers should be encouraged to make public their raw data so that they may be included in repositories such as Oncomine.

By applying this method to a near comprehensive collection of colorectal cancer expression profiling studies, we were able to determine the genes that were reported with a statistically significant frequency. As an extension of the previous version of the meta-analysis method, we categorized some genes according to their total number of tissue samples as lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3), instead of using absolute numbers. This allowed the average fold-change criteria to have a greater effect on the gene rank in cases where total sample sizes were similar. In the original version, fold-change rarely had any impact on rank.

4.2 Statistical Significance in Overlap in Two out of Three Comparisons

We observed that for the cancer versus normal and adenoma versus normal comparisons, genes were consistently reported as differentially expressed and

that the level of overlap was statistically significant. Conversely, while each of the five cancer versus adenoma studies reported differentially expressed genes, the level of overlap among these studies was not significant, suggesting that the number of multi-study genes in the five studies could have been observed due to chance. Determining the significance of overlap between studies provides another filtering step to remove false positive genes from further consideration. When ignoring the significance of the observed overlap, one may be misled by multi-study genes. For example, without knowledge of the statistical significance, one may reason that the multi-study genes in the cancer versus adenoma comparison are biologically relevant, even though this decision can not be reasonably made because the observed level of overlap may be due to chance alone.

4.3 Genes Reported with Inconsistent Direction of Differential Expression

In the cancer versus normal comparison, a total of 748 genes were reported as differentially expressed in at least two independent studies. While the majority of these genes were reported as differentially expressed in the same direction, 175 genes (23.4%) were not. Out of these 175 genes, 132 (75.4%) were reported in two studies, 32 (18.3%) were reported in three studies, eight (4.6%) were reported in four studies, two (1.1%) were reported in five studies, and one (0.6%) was reported in six studies. There are many potential explanations for these observed inconsistencies. Firstly, one limitation with such meta-analyses is the

overgeneralization of comparisons. While every effort was made to ensure that each study included in each of the three comparisons were comparable, there are bound to be inconsistencies due to the lack of relevant clinical data being reported in each of the studies. For example, in the cancer versus normal comparison, SLC26A3 was reported as down-regulated in five studies [5, 8-11], but as up-regulated in one study [12]. The five studies that reported this gene as down-regulated did not specify the microsatellite status of the colorectal cancer tissue samples being used, while the one study that reported the up-regulation of this gene utilized a mixture of microsatellite stable and unstable tissue samples. Other than microsatellite stability, other clinical features, such as the specific portion of the colon where the tissue samples were taken [9], may impact the direction of differential expression. Thus, due to the lack of this clinical data, it is difficult to determine whether the results of each independent study are truly comparable with each other. Conversely, if this clinical data was more readily available, more specific comparisons, such as microsatellite stable colorectal tissue samples taken from male patients versus paired normal mucosa, could be performed.

A related explanation for why some genes were reported as differentially expressed in an inconsistent direction is the heterogeneity in the tissue samples utilized. The independent studies experimented on tissue samples taken from vastly different populations, each with different genetic and environmental backgrounds that may contribute to differing expression profiles. Furthermore,

the tissue samples utilized by each study themselves will be heterogeneous compared to each other. In order to have adequate quantities of tissue to work with, most studies perform high throughput expression profiling on pooled tissue samples, which results in a gene expression signal that is 'averaged' across all cells in the samples [11]. However, the expression of a gene in a single cell may be drastically different from this average. Therefore, depending on how the tissue samples were isolated and which ones were pooled together, the genes may be reported as differentially expressed in an inconsistent direction. One of the studies included in the cancer versus normal comparison [11] investigated the feasibility of performing pooling tissue samples together by plotting the expression signal of all genes in a pooled sample versus the expression signal of genes from one of the samples in the pool. The authors calculated Pearson correlation coefficients and saw that their values ranged from 0.80 to 0.97, suggesting that the pooling of their specific tissue samples maintained patterns of gene expression representative of each distinct tissue sample. Such an analysis should be performed in studies utilizing pooled samples to ensure that the pooled versus unpooled results are comparable.

Finally, poor study design producing inaccurate results may also explain the presence of these genes. In many cases, these genes were ignored because one lone study reported an inconsistent direction of differential expression, which raises suspicions of the validity of the results of the lone disagreeing study. One concern is that some biologically relevant genes may be omitted due to such a

study. Therefore, it may be beneficial to include some genes where the majority the studies agreed on the direction of differential expression, instead of the much more stringent 'all-or-none' approach we have utilized. However, since the majority of these genes (75.4%) were reported in only two studies, including these genes would not alter the identity of the highest ranking candidates greatly.

Despite of these inconsistencies, we remind the reader that the majority of the multi-study genes (76.6%) were consistently reported as differentially expressed in the same direction, which is an encouraging result, given that each independent study utilized diverse experimental techniques and tissue samples.

4.4 Literature Review of Consistently Reported Differentially Expressed Genes

To further assess our results, we performed a literature review of the genes reported by at least seven studies in the cancer versus normal comparison to determine if any have been shown to have diagnostic and/or prognostic utility in colorectal cancer. The most consistently reported differentially expressed gene in our meta-analysis was carbonic anhydrase 2 (*ca2*), which was reported as down-regulated in 11 studies. Along with carbonic anhydrase 1 (*ca1*), *ca2* has been shown to have prognostic significance where the expression of the corresponding enzymes was related to the metastatic aggressiveness of colorectal cancer [13]. Similarly, the potential diagnostic utility of *ca2* was demonstrated in a study measuring CA2 protein levels in fecal matter. The

authors found that the average level of CA2 protein in the fecal matter of colorectal cancer patients was shown to be significantly greater than that which was found in the control group [14], which, interestingly, is inconsistent with the direction reported in the 11 studies. Immunohistochemistry (IHC) has been performed on colorectal tumor and healthy mucosa tissue to monitor the protein levels of four carbonic anhydrases, among them CA2 [15]. That study demonstrated the level of CA2 protein decreased in cancer relative to healthy tissue, thus confirming the transcript based expression profiling results.

Carbonic anhydrases catalyze the reversible reaction $CO_2 + H_20 \leftrightarrow H^+ + HCO_3^-$. To date, 16 isozymes have been identified in mammals [16]. In general, carbonic anydrases are divided into those that are localized in the cytoplasm (1, 2, 3, 5, 7, and 13) and those that are associated with the plasma membrane (4, 9, 12, 14, and 15) [16].

From our meta-analysis results, we observed that colorectal cancer cells show down-regulation of four carbonic anhydrases (1, 2, 4, and 12). However, we observed some inconsistencies between the results of published studies and that of the meta-analysis. For example, many studies have been conducted to show that the carbonic anhydrase 9 (CA9) enzyme is up-regulated in many different types of cancer, including colorectal [17, 18]. However, only one of the twenty-three cancer versus normal studies from our meta-analysis reported the up-regulation of *ca9* [12]. However, this discrepancy can potentially be explained by

the results of a recent study that compared the expression of CA9 at the protein level in HNPCC, MSS, and MSI colorectal samples [17]. This study concluded that CA9 was most strongly up-regulated in HNPCC. Since our meta-analysis focused on sporadic cancer which, as described earlier, makes up the majority of colorectal cancer cases, up-regulation of CA9 may be a unique property of inherited colorectal cancer cases.

The role of *ca9* in relation to cancer progression has been well studied. The Oncomine database reports that the expression of this gene has been studied in various types of cancer, such as bladder [19], lung [20], and ovarian [21]. CA9, a transmembrane carbonic anydrase with an extracellular active site, has been reported to promote the acidification of the extracellular environment [18]. This observation suggests that the forward reaction, in which CO_2 is hydrated, dominates over the reverse reaction. An acidic extracellular pH (pH_e) has been demonstrated to contribute to cancer progression by many different mechanisms [22, 23], such as by promoting normal cell death by inducing necrosis or caspase mediated activation of p53 dependent apoptosis pathways [24, 25]. It is expected that the death of normal cells produces more space into which cancer cells may proliferate [22]. An acidic pH_e also promotes angiogenesis through acid-induced release of vascular endothelial growth factor and interleukin-8 [26, 27]. Finally, an acidic pH_e has been shown to indirectly promote extracellular matrix degradation by inducing adjacent normal cells to release proteolytic enzymes, thus promoting cancer invasion [28].

Another inconsistency between the results of published studies and that of the meta-analysis involves carbonic anhydrase 12 (*ca12*). While its expression at the transcript level was reported as down-regulated in four independent studies in the cancer versus normal comparison, another study demonstrated, with IHC, its up-regulation at the protein level [15]. Currently, it is unclear how the *ca12* transcript appears to be down-regulated and yet an increased amount of CA12 protein is observed. Future studies should be conducted to determine if the half-lives of cancer and normal transcripts differ significantly. Finally, similar to CA9, CA12 is a transmembrane protein and has been reported to play a similar role in contributing to an acidic pH_e [18].

When considered together, the results from the published studies on carbonic anhydrases and from our meta-analysis suggest that colorectal cancer cells show up-regulation of some carbonic anhydrases, but down-regulation of others. While all known carbonic anhydrases catalyze the reversible reaction, they do so with varying catalytic efficiencies [16]. Furthermore, as described, the isozymes have different subcellular localizations. These two characteristics likely account for the different direction of differential expression of these isozymes. Perhaps CA9 and CA12 show up-regulation at the protein level because they are both membrane associated, which allows them to contribute to an acidic pH_e. Therefore, colorectal cancer cells may be selected for their ability to up-regulate these two isozymes. Conversely, CA1 and CA2, both of which were reported as

down-regulated by our meta-analysis, are localized in the cytoplasm of the cancer cell. Due to the hypoxic conditions of most cancer cells, anaerobic respiration is utilized [18, 23], which results in the production of energy, in the form of ATP, and lactic acid. This excess of lactic acid is exported out of the cell, along with H^+, to the extracellular space. Furthermore, H^+ is being exported by transporters, such as the Na^+-H^+ antiporter [18]. Therefore, to maintain homeostasis of the intracellular pH, CA1 and CA2 may favour catalyzing the forward reaction to produce H^+ and HCO_3^-. However, by having these two isozymes down regulated, there will potentially be more CO_2 that diffuses out of the cell, where CA9 and CA12 can catalyze the forward reaction to create an acidic pH_e. Thus, colorectal cancer cells may be selected for their ability to down-regulate CA1 and CA2. Therefore, the subcellular localization of an isozyme may impact its function and thus explain the observed direction of differential expression.

Given their potential roles in cancer progression, many carbonic anhydrase inhibitors have been developed. For example, derivatives of sulfonamides, such as methazolamide, currently used to treat glaucoma [29], acetzolamide, ethoxzolamide, indisulam, and others have been shown to be potent inhibitors of CA12 [29]. Indisulam, a compound that has also been shown to inhibit CA9 [29], is currently in clinical development as an antitumor drug [30, 31]. Thus, future studies should also consider the effectiveness of these inhibitors in colorectal cancer.

Transforming Growth Factor, beta-induced, 68 kDa (*tgfβi*) was reported as up-regulated in nine studies. TGFβI is a secreted extracellular matrix protein and was discovered through differential expression analysis of a transforming growth factor-β1 (TGF-β1)-treated human lung adenocarcinoma cell line [32, 33]. This gene has also been shown to be strongly induced by TGF-β1 in many other human cell lines [34, 35]. Despite of the consistent reporting of the over-expression of this gene, as far as we know, no study has focused specifically on its diagnostic and/or prognostic utility or its role in colorectal cancer tumorigenesis. Furthermore, over-expression at the protein level has yet to be validated with IHC.

Interferon induced transmembrane protein 1 (*ifitm1*) was also reported as up-regulated in nine studies. The protein product has been shown to mediate the antiproliferative properties of the interferon cytokines [36] and was observed to be over-expressed in gastric cancer cells, which resulted in tumor cells being more resistant to natural killer cells and produced a more invasive phenotype [37]. As far as we know, IHC on human colorectal cancer tissue have not been performed for IFITM1 protein, however, RT-PCR was previously conducted on adenomas in a murine model as well as a human colorectal carcinoma cell line, HT29, and elevated expression of *ifitm* genes (*ifitm1*, *ifitm2*, and *ifitm3*) was observed [38]. No further studies have considered the diagnostic and/or prognostic potential of *ifitm1* expression in colorectal cancer.

Mal, T-cell differentiation protein-like (*mall*), reported as down-regulated in seven studies, is a member of the MAL proteolipid family [39] and encodes an integral protein located in glycolipid- and cholesterol-enriched membranes. To the best of our knowledge, its expression at the protein level has not been measured by IHC and diagnostic and/or prognostic utilities have not been studied.

Carcinoembryonic antigen-related cell adhesion molecule 1 (*ceacam1*), reported as down-regulated in seven studies, has been shown to be a tumor suppressor in which expression is lost in adenomas and carcinomas. Moreover, the absence of *ceacam1* expression was shown to be correlated with reduced rates of apoptosis in polyps [40]. However, a retrospective study performed IHC on CEACAM1 and showed that 58% of colorectal cancer patients showed an increase in expression and that this increase was not related to overall survival or disease-free survival in colorectal cancer patients [41].

Secreted Protein, Acidic and Rich in Cysteine (*sparc*) was previously demonstrated by our group to be a putative resistance-reversal gene [42] and was reported as up-regulated in seven studies. Differentially expressed genes between resistant and sensitive human MIP101 colon cancer cells were determined and *sparc* was shown to be consistently down-regulated in the resistant cell lines. Their sensitivity was restored by re-expression of *sparc*, suggesting that its expression has prognostic utility. IHC performed on colorectal

cancer tissue samples showed increased staining of SPARC protein levels [43]. However, another IHC study [44] demonstrated down-regulation of SPARC due to methylation of its promoter. Further studies related to the role of *sparc* in colorectal tumorigenesis are currently underway in our group.

Growth Differentiation Factor 15 (*gdf15*), reported as up-regulated in seven studies, is a member of the TGFβ superfamily. Diagnostic and prognostic utility of *gdf15* in colorectal cancer has been suggested by studies that showed increased serum levels of GDF15 protein in colorectal cancer patients relative to healthy controls [45]. These levels increased during disease progression and may have clinical use in the management of colorectal cancer patients [46].

Myc, reported as up-regulated in seven studies in our meta-analysis, is a transcription factor that regulates various processes such as cell-cycle progression, differentiation, apoptosis, and cell motility [47]. IHC on MYC has shown that its expression increases during disease progression [48] and when combined with nuclear β-catenin expression, *myc* expression was demonstrated to have prognostic utility [49].

4.5 Consistently Reported Differentially Expressed Genes in both Cancer versus Normal and Adenoma versus Normal Comparisons

There were 26 differentially expressed genes that were reported with a statistically significant frequency in both the cancer versus normal and adenoma versus normal results. Up-regulated genes found in both comparisons consisted of those involved in ribosomal function and translation. It has been previously shown that up-regulation of the *myc* transcription factor induces genes encoding for ribosomal subunits and elongation factors [50]. This is to be expected, given the increased rates of protein synthesis that is needed to sustain the high proliferation rate of tumor cells. Conversely, down-regulated genes found in both comparisons included those involved in dehydrogenase activity and fatty acid oxidation. Elimination of fatty acid metabolic enzymes has been shown to be linked to reduced apoptosis in colonic epithelial cells [51].

Chapter 4 References

1. Shih W, Chetty R, Tsao MS: **Expression profiling by microarrays in colorectal cancer (review).** *Oncol Rep* 2005, **13:**517-524.

2. Cardoso J, Boer J, Morreau H, Fodde R: **Expression and genomic profiling of colorectal cancer.** *Biochim Biophys Acta* 2007, **1775:**103-137.

3. Sagynaliev E, Steinert R, Nestler G, Lippert H, Knoch M, Reymond MA: **Web-based data warehouse on gene expression in human colorectal cancer.** *Proteomics* 2005, **5:**3066-3078.

4. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM: **Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles.** *Neoplasia* 2007, **9:**166-180.

5. Notterman DA, Alon U, Sierk AJ, Levine AJ: **Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays.** *Cancer Res* 2001, **61:**3124-3130.

6. Zou TT, Selaru FM, Xu Y, Shustova V, Yin J, Mori Y, Shibata D, Sato F, Wang S, Olaru A, Deacu E, Liu TC, Abraham JM, Meltzer SJ: **Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon.** *Oncogene* 2002, **21:**4855-4862.

7. Griffith OL, Melck A, Jones SJ, Wiseman SM: **Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers.** *J Clin Oncol* 2006, **24:**5043-5051.

8. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y: **Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas.** *Oncogene* 2002, **21:**4120-4128.

9. Buckhaults P, Rago C, St Croix B, Romans KE, Saha S, Zhang L, Vogelstein B, Kinzler KW: **Secreted and cell surface genes expressed in benign and malignant colorectal tumors.** *Cancer Res* 2001, **61:**6996-7001.

10. Croner RS, Foertsch T, Brueckl WM, Guenther K, Siebenhaar R, Stremmel C, Matzel KE, Papadopoulos T, Kirchner T, Behrens J, Klein-Hitpass L, Stuerzl M, Hohenberger W, Reingruber B: **Common denominator genes that distinguish colorectal carcinoma from normal mucosa.** *Int J Colorectal Dis* 2005, **20:**353-362.

11. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ: **Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling.** *J Natl Cancer Inst* 2002, **94:**513-521.

12. Kim H, Nam SW, Rhee H, Shan Li L, Ju Kang H, Hye Koh K, Kyu Kim N, Song J, Tak-Bun Liu E, Kim H: **Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas.** *Oncogene* 2004, **23:**6218-6225.

13. Bekku S, Mochizuki H, Yamamoto T, Ueno H, Takayama E, Tadakuma T: **Expression of carbonic anhydrase I or II and correlation to clinical aspects of colorectal cancer.** *Hepatogastroenterology* 2000, **47:**998-1001.

14. Yokoyama S, Shatney CH, Mochizuki H, Hase K, Johnson DL, Cummings S, Trollope ML, Tamakuma S: **The potential role of fecal carbonic anhydrase II in screening for colorectal cancer.** *Am Surg* 1997, **63:**243-6; discussion 246-7.

15. Kivela AJ, Saarnio J, Karttunen TJ, Kivela J, Parkkila AK, Pastorekova S, Pastorek J, Waheed A, Sly WS, Parkkila TS, Rajaniemi H: **Differential expression of cytoplasmic carbonic anhydrases, CA I and II, and membrane-associated isozymes, CA IX and XII, in normal mucosa of large intestine and in colorectal tumors.** *Dig Dis Sci* 2001, **46:**2179-2186.

16. Esbaugh AJ, Tufts BL: **The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates.** *Respir Physiol Neurobiol* 2006, **154:**185-198.

17. Niemela AM, Hynninen P, Mecklin JP, Kuopio T, Kokko A, Aaltonen L, Parkkila AK, Pastorekova S, Pastorek J, Waheed A, Sly WS, Orntoft TF, Kruhoffer M, Haapasalo H, Parkkila S, Kivela AJ: **Carbonic anhydrase IX is highly expressed in hereditary nonpolyposis colorectal cancer.** *Cancer Epidemiol Biomarkers Prev* 2007, **16:**1760-1766.

18. Thiry A, Dogne JM, Masereel B, Supuran CT: **Targeting tumor-associated carbonic anhydrase IX in cancer therapy.** *Trends Pharmacol Sci* 2006, **27:**566-573.

19. Bredel M, Bredel C, Juric D, Harsh GR, Vogel H, Recht LD, Sikic BI: **Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas.** *Cancer Res* 2005, **65:**8679-8689.

20. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M: **Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.** *Proc Natl Acad Sci U S A* 2001, **98:**13790-13795.

21. Hendrix ND, Wu R, Kuick R, Schwartz DR, Fearon ER, Cho KR: **Fibroblast growth factor 9 has oncogenic activity and is a downstream target of wnt signaling in ovarian endometrioid adenocarcinomas.** *Cancer Res* 2006, **66:**1354-1362.

22. Gatenby RA, Gawlinski ET, Gmitro AF, Kaylor B, Gillies RJ: **Acid-mediated tumor invasion: A multidisciplinary study.** *Cancer Res* 2006, **66:**5216-5223.

23. Stubbs M, McSheehy PM, Griffiths JR, Bashford CL: **Causes and consequences of tumour acidity and implications for treatment.** *Mol Med Today* 2000, **6:**15-19.

24. Williams AC, Collard TJ, Paraskeva C: **An acidic environment leads to p53 dependent induction of apoptosis in human adenoma and carcinoma cell lines: Implications for clonal selection during colorectal carcinogenesis.** *Oncogene* 1999, **18:**3199-3204.

25. Park HJ, Lyons JC, Ohtsubo T, Song CW: **Acidic environment causes apoptosis by increasing caspase activity.** *Br J Cancer* 1999, **80:**1892-1897.

26. Shi Q, Le X, Wang B, Abbruzzese JL, Xiong Q, He Y, Xie K: **Regulation of vascular endothelial growth factor expression by acidosis in human cancer cells.** *Oncogene* 2001, **20:**3751-3756.

27. Fukumura D, Xu L, Chen Y, Gohongi T, Seed B, Jain RK: **Hypoxia and acidosis independently up-regulate vascular endothelial growth factor transcription in brain tumors in vivo.** *Cancer Res* 2001, **61:**6020-6024.

28. Rozhin J, Sameni M, Ziegler G, Sloane BF: **Pericellular pH affects distribution and secretion of cathepsin B in malignant cells.** *Cancer Res* 1994, **54:**6517-6525.

29. Vullo D, Innocenti A, Nishimori I, Pastorek J, Scozzafava A, Pastorekova S, Supuran CT: **Carbonic anhydrase inhibitors. inhibition of the transmembrane isozyme XII with sulfonamides-a new target for the design of antitumor and antiglaucoma drugs?** *Bioorg Med Chem Lett* 2005, **15:**963-969.

30. Brown JM, Wilson WR: **Exploiting tumour hypoxia in cancer treatment.** *Nat Rev Cancer* 2004, **4:**437-447.

31. Scozzafava A, Owa T, Mastrolorenzo A, Supuran CT: **Anticancer and antiviral sulfonamides.** *Curr Med Chem* 2003, **10:**925-953.

32. Schneider D, Kleeff J, Berberat PO, Zhu Z, Korc M, Friess H, Buchler MW: **Induction and expression of betaig-h3 in pancreatic cancer cells.** *Biochim Biophys Acta* 2002, **1588:**1-6.

33. Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD, Purchio AF: **cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta.** *DNA Cell Biol* 1992, **11:**511-522.

34. Skonier J, Bennett K, Rothwell V, Kosowski S, Plowman G, Wallace P, Edelhoff S, Disteche C, Neubauer M, Marquardt H: **Beta ig-h3: A transforming growth factor-beta-responsive gene encoding a secreted protein that inhibits cell attachment in vitro and suppresses the growth of CHO cells in nude mice.** *DNA Cell Biol* 1994, **13:**571-584.

35. LeBaron RG, Bezverkov KI, Zimber MP, Pavelec R, Skonier J, Purchio AF: **Beta IG-H3, a novel secretory protein inducible by transforming growth factor-beta, is present in normal skin and promotes the adhesion and spreading of dermal fibroblasts in vitro.** *J Invest Dermatol* 1995, **104:**844-849.

36. Akyerli CB, Beksac M, Holko M, Frevel M, Dalva K, Ozbek U, Soydan E, Ozcan M, Ozet G, Ilhan O, Gurman G, Akan H, Williams BR, Ozcelik T: **Expression of IFITM1 in chronic myeloid leukemia patients.** *Leuk Res* 2005, **29:**283-286.

37. Yang Y, Lee JH, Kim KY, Song HK, Kim JK, Yoon SR, Cho D, Song KS, Lee YH, Choi I: **The interferon-inducible 9-27 gene modulates the susceptibility to natural killer cells and the invasiveness of gastric cancer cells.** *Cancer Lett* 2005, **221:**191-200.

38. Andreu P, Colnot S, Godard C, Laurent-Puig P, Lamarque D, Kahn A, Perret C, Romagnolo B: **Identification of the IFITM family as a new molecular marker in human colorectal tumors.** *Cancer Res* 2006, **66:**1949-1955.

39. de Marco MC, Kremer L, Albar JP, Martinez-Menarguez JA, Ballesta J, Garcia-Lopez MA, Marazuela M, Puertollano R, Alonso MA: **BENE, a novel raft-associated protein of the MAL proteolipid family, interacts with caveolin-1 in human endothelial-like ECV304 cells.** *J Biol Chem* 2001, **276:**23009-23017.

40. Nittka S, Gunther J, Ebisch C, Erbersdobler A, Neumaier M: **The human tumor suppressor CEACAM1 modulates apoptosis and is implicated in early colorectal tumorigenesis.** *Oncogene* 2004, **23:**9306-9313.

41. Jantscheff P, Terracciano L, Lowy A, Glatz-Krieger K, Grunert F, Micheel B, Brummer J, Laffer U, Metzger U, Herrmann R, Rochlitz C: **Expression of CEACAM6 in resectable colorectal cancer: A factor of independent prognostic significance.** *J Clin Oncol* 2003, **21:**3638-3646.

42. Tai IT, Dai M, Owen DA, Chen LB: **Genome-wide expression analysis of therapy-resistant tumors reveals SPARC as a novel target for cancer therapy.** *J Clin Invest* 2005, **115:**1492-1502.

43. Lussier C, Sodek J, Beaulieu JF: **Expression of SPARC/osteonectin/BM4O in the human gut: Predominance in the stroma of the remodeling distal intestine.** *J Cell Biochem* 2001, **81:**463-476.

44. Yang E, Kang HJ, Koh KH, Rhee H, Kim NK, Kim H: **Frequent inactivation of SPARC by promoter hypermethylation in colon cancers.** *Int J Cancer* 2007, **121:**567-575.

45. Welsh JB, Sapinoso LM, Kern SG, Brown DA, Liu T, Bauskin AR, Ward RL, Hawkins NJ, Quinn DI, Russell PJ, Sutherland RL, Breit SN, Moskaluk CA, Frierson HF,Jr, Hampton GM: **Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum.** *Proc Natl Acad Sci U S A* 2003, **100:**3410-3415.

46. Brown DA, Ward RL, Buckhaults P, Liu T, Romans KE, Hawkins NJ, Bauskin AR, Kinzler KW, Vogelstein B, Breit SN: **MIC-1 serum level and genotype: Associations with progress and prognosis of colorectal carcinoma.** *Clin Cancer Res* 2003, **9:**2642-2650.

47. Vita M, Henriksson M: **The myc oncoprotein as a therapeutic target for human cancer.** *Semin Cancer Biol* 2006, **16:**318-330.

48. Xie D, Sham JS, Zeng WF, Lin HL, Che LH, Wu HX, Wen JM, Fang Y, Hu L, Guan XY: **Heterogeneous expression and association of beta-catenin, p16 and c-myc in multistage colorectal tumorigenesis and progression detected by tissue microarray.** *Int J Cancer* 2003, **107:**896-902.

49. Bondi J, Bukholm G, Nesland JM, Bukholm IR: **Expression of non-membranous beta-catenin and gamma-catenin, c-myc and cyclin D1 in relation to patient outcome in human colon adenocarcinomas.** *APMIS* 2004, **112:**49-56.

50. Kim S, Li Q, Dang CV, Lee LA: **Induction of ribosomal genes and hepatocyte hypertrophy by adenovirus-mediated expression of c-myc in vivo.** *Proc Natl Acad Sci U S A* 2000, **97:**11198-11202.

51. Augenlicht LH, Anthony GM, Church TL, Edelmann W, Kucherlapati R, Yang K, Lipkin M, Heerdt BG: **Short-chain fatty acid metabolism, apoptosis, and apc-initiated tumorigenesis in the mouse gastrointestinal mucosa.** *Cancer Res* 1999, **59:**6005-6009.

Chapter 5: Conclusion and Future Studies

The results of this thesis research have demonstrated that the application of the meta-analysis method that was developed can identify consistently reported differentially expressed genes in colorectal cancer. That some of the high ranking candidates have already been shown to have diagnostic and/or prognostic utility in colorectal cancer is a very encouraging result and they collectively act as a positive control.

5.1 Validation of Transcript based Gene Expression Profiling Results on Colorectal Tissue Microarrays

Tissue microarrays have emerged as a powerful tool to validate the candidate biomarkers identified in transcript based expression profiling studies [1]. They consist of a glass slide that contains tissue samples arranged in an array format. Construction of tissue microarrays begin with collection of archival paraffin-embedded, formalin-fixed tissues. Next, representative areas from each tissue are selected by a pathologist. Tissue cores, approximately 0.6 to 2.0 mm in diameter, are removed and then transferred to a recipient block. The resulting tissue microarray is cut into sections and then transferred to glass slides. At this point, IHC with fluorescently labeled antibodies can be performed to quantify the expression levels of the proteins of interest [2, 3]. The intensity and localization of the antibody signal within the tissues are examined by pathologists and scored based on a predefined semi-quantitative scale.

One area of future work would be to have a portion of these consistently reported differentially expressed genes validated on tissue microarrays. Antibodies specific to 21 of the high ranking candidates have been ordered and screening the protein expression of those candidates will begin shortly. These candidates were chosen based on their ranking, antibody availability, and their relationship to progression in various types of cancer. Another area of future work involves determining whether the protein expression of the candidates can subdivide patients into prognostic subgroups using various clustering and classification methods. Clinical data from the patients in which the tissue microarrays were generated will be compiled.

In conclusion, the results of this meta-analysis identified genes already shown to have diagnostic and/or prognostic potential in colorectal cancer. Perhaps more interesting are the genes, such as *tgfβi* and *ifitm1*, that were consistently reported but have yet to be studied specifically as biomarkers. Also, the genes further down the list (i.e., those identified as differentially expressed by four, five, six, etc. independent studies) warrant further investigation. Further studies focused on these genes will aid in determining a panel of diagnostic and prognostic colorectal cancer biomarkers with sufficient sensitivity and specificity.

Chapter 5 References

1. Giltnane JM, Rimm DL: **Technology insight: Identification of biomarkers with tissue microarray technology.** *Nat Clin Pract Oncol* 2004, **1:**104-111.

2. Wan WH, Fortuna MB, Furmanski P: **A rapid and efficient method for testing immunohistochemical reactivity of monoclonal antibodies against multiple tissue samples simultaneously.** *J Immunol Methods* 1987, **103:**121-129.

3. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens.** *Nat Med* 1998, **4:**844-847.