ALTERNATIVE POLYADENYLATION ANALYSIS IN MYELODYSPLASTIC SYNDROMES, GLIOBLASTOMA AND GASTRIC CANCER

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

TUNA ÇINKILLI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

DECEMBER 2014

Approval of the thesis:

ALTERNATIVE POLYADENYLATION ANALYSIS IN MYELODYSPLASTIC SYNDROMES, GLIOBLASTOMA AND GASTRIC CANCER

submitted by **TUNA ÇINKILLI** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Orhan Adalı _____
Head of Department, **Biology**

Assoc. Prof. Dr. A. Elif Erson-Bensan _____
Supervisor, **Biology Dept., METU**

Examining Committee Members:

Assoc. Prof. Dr. Sreeparna Banerjee _____
Biology Dept., METU

Assoc. Prof. Dr. A. Elif Erson-Bensan _____
Biology Dept., METU

Assoc. Prof. Dr. Tolga Can _____
Computer Engineering Dept., METU

Assoc. Prof. Dr. Mesut Muyan _____
Biology Dept., METU

Assoc. Prof. Dr. Özlen Konu _____
Molecular Biology and Genetics Dept., Bilkent University

Date: 24.12.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Tuna Çınkıllı

Signature:

ABSTRACT

ALTERNATIVE POLYADENYLATION ANALYSIS IN MYELODYSPLASTIC SYNDROMES, GLIOBLASTOMA AND GASTRIC CANCER

Çınkıllı, Tuna
M.S., Department of Biology
Supervisor: Assoc. Prof. Dr. A. Elif Erson-Bensan

December 2014, 149 pages

Alternative polyadenylation (APA) is the process by which different length transcript isoforms are generated. Upon the changes in poly(A) tail positioning, especially in the 3' untranslated regions (UTRs), resulting mRNA isoforms may be regulated differently by negative or positive regulatory elements. Under normal circumstances, differential APA patterns have been observed during development or proliferation events. Therefore, such differential preferences of polyadenylation sites may be important in proliferative diseases. Indeed, APA events appear to be altered in cancer cells. A general trend towards 3' UTR shortening was observed in several cancers. The main aim of this study was to investigate alterations in APA patterns of myelodysplastic syndromes, glioblastoma multiforme and gastric cancer which are aggressive, rapidly proliferating diseases. To this end, APADetect, a probe-based microarray analysis tool, was used to analyze microarray datasets. For each transcript, the short/long ratio (SLR) values were calculated using the proximal and distal probe intensities according to poly(A) site location. Significant alterations in the SLR values were then detected by Significance Analysis of Microarrays (SAM).

Further analyses were carried out to reveal any potential correlations with patients characteristics. Significantly altered transcripts were also analyzed for functional enrichment using gene ontology tools. Indeed, enrichment patterns could be seen in transcripts with functions related to both cell division cycle and tissue specificity. In addition to the bioinformatics based analyses, we focused on the *TCF3* gene transcript which we observed to be commonly altered in the three diverse types of proliferative diseases that we analyzed and experimentally confirmed the presence of the shorter 3' UTR transcript isoform in a gastric cancer cell line. These results indicate that changes in APA patterns may be playing important roles in malignant transformations. While the results reported here are preliminary, we plan to extend these observations to an experimental setup to verify and investigate functional relevance. Further studies are necessary to identify novel diagnostic or prognostic markers for use in the medical field.

Keywords: Alternative Polyadenylation, 3’UTR, Myelodysplastic Syndromes, Glioblastoma Multiforme, Gastric Cancer

ÖZ

MYELODİSPLASTİK SENDROMLAR, GLİYOBLASTOM VE MİDE KANSERİNDE ALTERNATİF POLİADENİLASYON ANALİZİ

Çınkıllı, Tuna
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Doç. Dr. A. Elif Erson-Bensan

Aralık 2014, 149 sayfa

Alternatif poliadenilasyon (APA) farklı uzunluklara sahip transkript izoformlarını oluşturan mekanizmadır. Özellikle 3' UTR (transle olmayan bölge) değişikliklerine bağlı olarak oluşan mRNA izoformları negatif veya pozitif düzenleyici elementler tarafından farklı düzenlemelere uğrayabilmektedir. Normal şartlarda, gelişme veya hücre bölünmesi gibi olaylarda özelleşmiş APA tercihleri görülebilir. Bu nedenlerle APA değişiklikleri yüksek bölünme özelliği gösteren hastalıklarda önemli olabilir. Gerçekten de APA olayları kanser hücrelerinde önemli görünmektedir. Farklı kanser türlerinde, 3'UTR kısalmasına yönelik bir eğilim görülmüştür. Bu çalışmanın genel amacı, agresif ve hızlı bölünme gösteren hastalıklar olan myelodisplastik sendromlar, gliyoblatom ve mide kanserlerinde APA değişikliklerini araştırmaktır. Bu amaçla, mikroçip datasetlerini analiz etmek için, prob-tabanlı mikroçip analiz aracı olan APADetect kullanıldı. Her transkript için, poly(A) eklenme bölgesine bağlı olarak proksimal ve distal prob sinyallerinden kısa/uzun oranı (SLR) hesaplandı. Kısa/uzun oranında önemli değişiklikler Mikroçiplerin Anlamlılık Analizi (SAM) ile belirlendi.

Daha sonra yapılan analizlerle bu değişikliklerin hasta özellikleri ile herhangi bir korelasyon gösterip göstermediğine bakıldı. Önemli derecede değişen transkriptler ayrıca gen ontolojisi araçları kullanılarak görevsel yoğunlaşmalar için analiz edildi. Gerçekten de özellikle hücre bölünme döngüsünde veya dokuya özel görevleri olan birtakım genlerde yoğunlaşma eğilimleri görüldü. Biyoinformatik analizlere ek olarak, incelediğimiz üç farklı proliferatif hastalıkta ortak olarak değişen *TCF3* gen transkriptiyle ilgilendik ve kısa 3' UTR izoformun varlığını bir mide kanseri hücre hattında deneysel olarak doğruladık. Bu sonuçlar APA motiflerindeki değişikliklerin habis dönüşümlerde önemli roller oynayabileceğini göstermektedir. Burada bildirilen sonuçlar başlangıç niteliğinde olmakla birlikte, işlevsel ilişkilendirmeleri doğrulamak ve araştırmak için gözlemlediklerimizi deneysel kurulumla genişletmeyi amaçlıyoruz. Yeni tanısal veya prognostik markerlerin tıp alanında kullanılmak üzere belirlenmesi için daha ileri çalışmalar gerekmektedir.

Anahtar Kelimeler: Alternatif Poliadenilasyon, 3’UTR, Myelodisplastik Sendromlar, Gliyoblastom, Mide Kanseri

To my family

ACKNOWLEDGEMENTS

First and foremost, I wish to express my deepest and sincere gratitude to my supervisor Assoc. Prof. Dr. A. Elif Erson-Bensan for finding me worthy to be her pupil, for her endless support, patience, and knowledge. I have admired and respected her deeply ever since I was an undergraduate and she has been an inspiration to me long before she even knew.

I would like to thank all my thesis committee members; Assoc. Prof. Dr. Sreeparna Banerjee, Assoc. Prof. Dr. Tolga Can, Assoc. Prof. Dr. Mesut Muyan, and Assoc. Prof. Dr. Özlen Konu.

I would like to express my sincere gratitude to Assoc. Prof. Dr. Tolga Can for the bioinformatics support. I am deeply grateful to Assoc. Prof. Dr. Mesut Muyan for his invaluable suggestions and inspiration.

I wish to thank all members of Erson Lab; Taner Tuncer, Begüm Akman, Hızlan Hıncal Ağuş, Nihan Özdemirler, Merve Öyken, Ülkü Uzun, Esra Yavuz and Gülten Tuncel for tirelessly answering my never-ending questions and for their much appreciated friendships. This would not have been possible without them. I wish to thank Muyan Lab members; Gizem Güpür, Pelin Yaşar, Sırma Damla User, Gamze Ayaz, and Hasan Hüseyin Kazan who have become good friends and never hesitated to help in my time of need. I wish to thank Banerjee Lab for sharing their resources.

I am truly thankful to my entire family, who raised me to become who I am today. My deepest thanks go to my sisters Evren Bıçakçı, Zeynep Çınkıllı and Işık Çay. They are my angels, the three suns that continuously shine and illuminate my life simply by existing.

I cannot express my thanks enough to my best friends who have been through it all with me. I cannot give all their names however much I may wish to, but I love them all the same. Each one of them has been invaluable throughout my life. Many special

thanks to my dear friends Ayşen Biten and Ozan Alican who have been by my side through thick and thin for nine years and for many more years to come.

I wish to thank TUBITAK for supporting me by National Scholarship Program for MSc Students (2012-2014).

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

CHAPTER 1

INTRODUCTION

1.1 Alternative Polyadenylation

1.1.1 3' end Formation of Mature mRNA

RNA processing is an important regulatory mechanism of gene expression. Processing events include capping, editing, splicing, and polyadenylation. Eukaryotic mRNAs are transcribed by RNA polymerase II and almost all mRNAs contain poly(A) tails [1,2]. Polyadenylation occurs in a reaction of two steps that include an endonucleolytic cleavage followed by poly(A) tail addition [3]. Polyadenylation factors recognize specific sites on precursor mRNA (pre-mRNA) for endonucleolytic cleavage and the poly(A) tail is added by poly(A) polymerases (PAPs) which are independent of template [4]. The average poly(A) tail length in humans is between 250 and 300 adenines [5].

Alternative polyadenylation is the process by which the position of poly(A) tail addition is chosen and different length transcripts are generated in the cell [6]. Proximal poly(A) site usage indicates formation of shorter transcripts, whereas longer transcript isoform is generated when distal poly(A) site is chosen [7].

3' UTR and poly(A) tail lengths of mRNAs affect properties of the transcripts such as mRNA half-life, transport to cytoplasm, or translation to protein [8]. Hence, the processing of mRNA 3' ends is an important gene expression regulation mechanism, and it will directly affect protein levels in the cell. Therefore, it is important to know about the mechanisms underlying these events and to observe how alterations in these pathways contribute to disease pathologies such as in cancer.

1.1.2 Process of Polyadenylation

As mentioned before, polyadenylation requires cleavage of pre-mRNA 3' end, followed by the addition of a poly(A) tail. This two-step reaction involves the activity of four multi-subunit protein complexes which are CstF (cleavage stimulation factor), CPSF (cleavage and polyadenylation specificity factor), CFI (cleavage factor I), and CFII (cleavage factor II). Processing is mediated by poly(A) polymerase (PAP) and RNA polymerase II (RNAP II) as well [9].

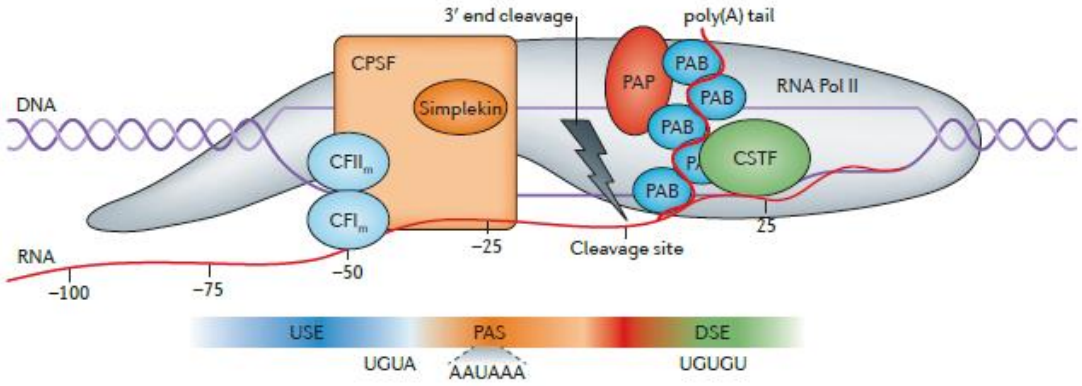


Figure 1.1 Core cleavage and polyadenylation factors. CPSF, CstF, CFI and CFII complexes perform cleavage. PAP and poly(A)-binding proteins (PABs) both assist cleavage and perform polyadenylation.

Cleavage, enhanced by upstream and downstream sequence elements (USE/DSE), occurs at poly(A) signal (PAS) position. Figure is adapted from [5].

The entire polyadenylation complex is recruited to the pre-mRNA as CPSF and CstF protein complexes interact with poly(A) signals simultaneously. Then the two-step polyadenylation reaction occurs for the completion of 3' end formation of the mRNA in a transcription-coupled manner.

1.1.2.1 Polyadenylation Types

In recent years, it has become evident that alternative polyadenylation is a key regulatory mechanism in gene expression, and over 50% of human genes are subject to this regulation [10].

mRNAs containing only one polyadenylation signal (PAS) in their 3' UTRs will always be polyadenylated from the same site. mRNAs containing more than one PAS will be adenylated from either proximal or distal PAS positions on 3' UTR, resulting in formation of different length isoforms. Not only the 3' UTR, but also the coding regions of mRNAs may contain multiple PASs both in intronic and exonic sequences. As a result of APA in the coding region, both different length mRNA isoforms and different protein isoforms will be generated in the cell [2]. These events are illustrated in Figure 1.2.

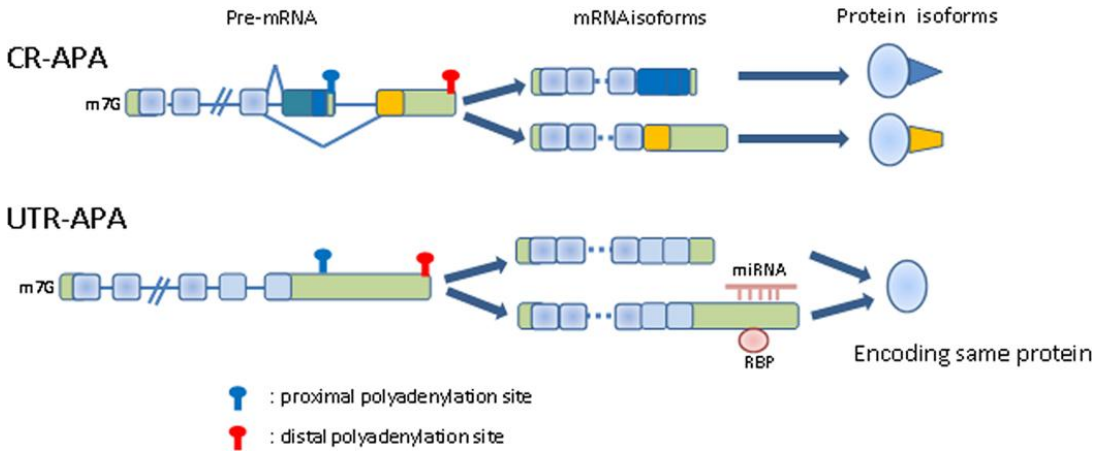


Figure 1.2 Alternative polyadenylation types. APA can occur inside the coding region or the 3' UTR of pre-mRNAs. When it occurs in the coding region (CR-APA), resulting mRNA is translated into different protein isoforms and the effect is qualitative. On the other hand, when it occurs in the 3' UTR (UTR-APA), translated protein levels are affected quantitatively due to loss of binding sites for regulatory elements such as microRNAs (miRNAs) or RNA-binding proteins (RBPs). Figure is taken from [9].

1.1.2.2 APA Regulation at the Molecular Level

APA occurs under regulation of *cis*-acting RNA elements, trans-acting regulatory proteins and epigenetic effector mechanisms.

1.1.2.2.1 *Cis*-acting Regulators

One core *cis*-acting RNA element is the polyadenylation signal (PAS). PASs are conserved hexameric sequences that direct the position of cleavage and polyadenylation [11]. The most commonly used conserved PAS is AAUAAA sequence that is positioned

20-30 nucleotides upstream to the cleavage site. However, recent studies have shown the presence and usage of alternate PASs located on the mRNAs [7].

In addition, other regulatory sequence motifs located upstream and downstream of PAS have been identified. Two upstream sequence motifs are the TTTTTTTT motif and AAWAAA motif (W standing for T or A). On the other hand, the palindromic CCAGSCTGG motif (S standing for G or C) is located downstream from PAS. These sequences have been identified near PASs that are seen in intra- or intergenic sequences [12].

Finally, 3' UTRs of mRNAs may fold into specific secondary structures or stem-loops [13]. Formation of such structures could potentially increase or downregulate the activities of polyadenylation regulating proteins [9].

1.1.2.2.2 Trans-acting Regulators

Trans-acting regulators of APA include a large variety of proteins that mediate the choice of PAS to be used as cleavage site on the mRNAs. Both 3'-processing proteins and RNA-binding proteins (RBPs) have been shown to affect APA [9]. These trans-acting factors may exert their effects in tissue-specific manners or not.

For 3'-processing proteins, the levels of CstF and CFI complex factor subunits are important in the choice of PAS to be used. The increased or decreased levels are associated with enhanced or repressed binding affinity of the proteins, respectively [14,15].

For RBPs, there are different examples of proteins inhibiting or enhancing the proximal poly(A) signal usage [9]. Distal poly(A) site usage enhancer proteins include *NOVA2* (neuro-oncological ventral antigen 2), *PABPN1* (polyadenylation binding protein nuclear 1), and *PTB1* (polypyrimidine-tract-binding protein 1) [16-18]. A proximal poly(A) site usage enhancing example is CPEB1 [19].

1.1.2.2.3 Epigenetic Regulation

It has recently been shown that the epigenetic modifications and chromatin structure are important in the choice of polyadenylation site. Poly(A) sites, being A-T rich, are depleted of nucleosomes conforming to a more open structure. In contrast, nucleosomes have been found to be enriched at downstream regions of poly(A) sites. As a result of this, the more open poly(A) site becomes more preferable to protein binding and recruitment of processing factors [20].

Not only sequence specific nucleosome depletions, but also histone modifications affect the chromatin structure. Hence, the histone code around the poly(A) site may also be important in regulating APA [21].

1.1.3 APA Interactions with Transcription and mRNA Processing

1.1.3.1 Transcription and APA

As the 3' end processing of mRNAs is thought to be coupled to transcription, it is interesting to investigate the extent of the relationship between APA and transcription [22]. There are two models by which transcription and APA are thought to be linked.

The first model suggests that proximal poly(A) site usage increases as transcription and processing machineries interact with each other closely [9]. This model's examples include PAF1c (RNA polymerase II-associated factor complex) and ELL2 proteins that have roles in both transcription and 3' end processing [23,24], both of which favor the proximal poly(A) site usage.

The second model is related to kinetics of transcription. As polyadenylation is transcription coupled, the proximal PASs will be exposed to 3' end processing factors long before the distal PASs especially if the rate of transcription is slower [25,26]. Hence,

slower elongation rate may cause proximal poly(A) sites to be favored, leading to shorter 3' UTRs.

1.1.3.2 Splicing and APA

APA is coupled to not only transcription, but also splicing. Splicing and polyadenylation machinery factors may interact with each other resulting in increased cleavage efficiency, especially during the events concerning the terminal exon or intronic sequences of pre-mRNAs [27].

Known examples of such interactions include proteins with roles in APA as well as splicing, such as Nova2 whose knockdown results in shortened 3' UTRs [16], CPEB1 which can recruit polyadenylation machinery and thereby hinder splicing mechanism's activity [19] and U1 whose knockdown results in choice of intronal polyadenylation site [28].

1.1.4 Detection of APA Events on a Global Scale

In recent years, APA has been found to be a highly used mechanism in control of gene expression. Today's technology offers different bioinformatics tools and databases that can be used in genome-wide detection of APA.

One of the earlier tools is expressed sequence tag (EST) databases that present data that is available for poly(A) site searches. EST sequence analyses have aided the revelation of numerous novel poly(A) sites [29].

Microarrays make use of mRNA-specific probes to detect genome-wide changes in gene expression patterns [30] and provide a powerful tool that can be used in APA analyses as well. The limitation with microarrays is that specific probe positions narrow down the analyses to known poly(A) sites; hence, novel polyadenylation events may not

be detected. Probe-based analysis of microarray data can still provide novel information on APA regulated genes.

More recently, second generation sequencing technologies have become available for transcriptome analyses and RNA-Seq technology was generated [31]. Analysis of RNA-Seq data can provide extensive information related to unidentified poly(A) sites available in the transcriptome as well as expression of alternatively polyadenylated transcript isoforms [32]. However powerful and 3'-end modified RNA-Seq is needed for identifying APA events in the transccriptome.

Finally, deep-sequencing based approaches derived from the design of RNA-Seq have been developed for better analyses of APA events on the transcriptomic scale. This is especially valid for transcripts that are not limited to containing 1-2 polyadenylation sites and instead three or more poly(A) sites dictate the transcript isoforms' generation in the cells [12]. Accordingly, statistical methods also require adaptations to such methods for more accurate evaluation of the data at hand [33].

To summarize, more extensive approaches are being developed for the analysis of APA events on a global scale to better understand this level of gene expression regulation.

1.1.5 APA Importance in Living Systems

APA has been shown to be an important regulatory mechanism of gene expression and it can occur in tissue specific manners. Examples of such commonly seen events include, 3' UTR shortening upon mouse B cell activation [34] or in proliferating T cells [35]. In other cases, 3' UTR lengthening may also occur as shown in mouse C2C12 cell differentiation indicating both tissue and development specific APA regulation [36].

The major consequence of 3' UTR length alterations is that 3' UTRs contain binding sites for regulatory elements such as microRNAs (miRNAs) which usually lead to suppression of target mRNA expression [37]. Considering that miRNA expression and APA profiles are observed to be altered significantly in cancer, it is not surprising that

these events contribute to tumorigenecity in organisms. Hence, further studies regarding alternative polyadenylation and its relation with different cancers are necessary to better understand the extremely complex nature of malignant proliferative diseases.

The focus of this study were APA event changes in myelodysplastic syndromes, glioblastoma multiforme and gastric cancer. These diverse diseases were of interest because they are highly malignant. Diagnosis is difficult and prognosis prediction is problematic. They all show low survival rates. Therefore, APA event changes were investigated to better explain the pathology of these three diverse types of proliferative diseases.

1.2 Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS) are a group of hematologic diseases that usually manifest as drug-resistant anemia due to ineffective hematopoiesis, and abnormal myeloid cell differentiation. MDSs are also associated with a risk of progression into acute myeloid leukemia [38]. They are the most common blood malignancies seen in patients of older ages [39]. Although some well-defined common alterations in the genetic structure are observed, about half of the patients have normal karyotypes indicating that chromosomal abnormalities are not the only mechanism or risk factor for MDSs and their progression into AML [40].

The diagnosis of MDSs requires bone marrow biopsies [39] and examination of $CD34^+$ hematopoietic stem cells (HSCs) [41]; however, it is not only costly but also, a very difficult test to apply especially considering the old age of patients that commonly suffer from these syndromes [39]. In addition, classification of MDSs into distinct subtypes can be highly subjective and is a major challenge in correct diagnosis, as well as in choice of therapy. Current classification of MDSs has been standardized by the World Health Organization (WHO) in 2008 [41]. The details of classification are given in Table 1.1.

Table 1.1 Current classification of myelodysplastic syndromes. Table is adapted from [41].

Disease Subtype	Blood Characteristics	Bone Marrow Characteristics
Refractory cytopenia with unilineage dysplasia (RCUD): (refractory anemia [RA]; refractory neutropenia [RN]; refractory thrombocytopenia [RT])	Unicytopenia or bicytopenia No or rare blasts* (< 1%)	Unilineage dysplasia < 5% blasts < 15% of erythroid precursors are ring sideroblasts**
Refractory anemia with ring sideroblasts (RARS)	Anemia No blasts	≥ 15% of erythroid precursors are ring sideroblasts Erythroid dysplasia only < 5% blasts
Refractory cytopenia with multilineage dysplasia (RCMD)	Cytopenia(s) No or rare blasts (< 1%) No Auer rods***	Dysplasia in ≥ 2 myeloid lineages < 5% blasts in marrow No Auer rods ± 15% ring sideroblasts
Refractory anemia with excess blasts-1 (RAEB-1)	Cytopenia(s) < 5% blasts No Auer rods	Uni- or multilineage dysplasia 5%-9% blasts No Auer rods
Refractory anemia with excess blasts-2 (RAEB-2)	Cytopenia(s) 5%-19% blasts Auer rods ±	Uni- or multilineage dysplasia 10%-19% blasts Auer rods ±
Myelodysplastic syndrome—unclassified (MDS-U)	Cytopenias < 1% blasts	Unequivocal dysplasia in one or more myeloid lineages < 5% blasts
MDS associated with isolated del(5q)	Anemia Usually normal or increased platelet count No or rare blasts (< 1%)	< 5% blasts Isolated del(5q) cytogenetic abnormality No Auer rods

*Blasts: Immature precursor cells

**Ring sideroblasts: Abnormal structures around immature erythrocyte nuclei due to impairment of iron incorporation

***Auer rods: Abnormal, needle-shaped inclusions in leukemic blasts

Survival estimation of MDS patients constitutes another issue. Prediction of prognosis cannot be performed easily, because overall survival depends on several different parameters including specific genomic alterations observed in patients, patient responses to therapeutic choices [38].

In light of all these information, alterations in APA profiles of normal karyotype MDS patients as compared to normal hematopoietic stem cells are of interest in the hopes that identification of novel diagnostic or prognostic indicators may potentially be possible.

1.3 Glioblastoma Multiforme

Gliomas are the most common and highly malignant brain tumors arising from the glial cells of the brain. They are classified into many subclasses according to the glial cell types from which they arise, and tumor grade. Among these subclasses, glioblastoma multiforme (grade IV), also called GBM, is the most commonly observed subtype that starts from specific glial cells that are called astrocytes [42]. Despite numerous and diverse therapy options (surgical removal, radiotherapy, chemotherapy, etc.), owing to the highly invasive nature of GBM, in most cases recurrence of tumor is observed resulting in a median survival of only 14.6 months. Most cases of GBM are primary tumors that occur *de novo*, whereas a small portion of patients exhibit GBM secondary to a lower grade astrocytic tumor [43].

Not surprisingly, GBM constitutes a group of heterogeneous tumors that can be classified according to their genetic alterations or expression profiles [43]. Considering the aggressive nature of these tumors, it is important to investigate and learn more about the molecular mechanisms underlying this disease for development of more effective therapies in order to increase survival rates. In this regard, it is known that under normal circumstances, mRNAs transcribed in the spinal cord and the brain tend to have longer 3' UTRs [44]. Hence, investigation of potential changes in APA patterns of GBM tumor tissues as compared to normal brain tissues is of interest and will contribute to the knowledge about gene expression alterations in cancer tissues.

1.4 Gastric Cancer

Gastric cancer is cancer of the stomach. It is the fourth most frequent cancer globally [45]. Its incidence is high and it shows poor survival. That is mostly due to the inadequacy of detection in the early stages of the disease, resulting commonly in metastasis. Hence, curative surgery, although effective in some cases, is not an applicable choice of treatment for most patients [46]. Among the risk factors of gastric cancer; age, gender, smoking, diet and *H. pylori* infection have all been included [47].

As seen in most cancers, gastric cancer is a heterogeneous disease and different classifications are made. Histologically, they are grouped according to the cells' differentiation status (differentiated or undifferentiated) or the type of glandular structure (intestinal or diffuse types). Different classes of tumors have been shown to exhibit differential expression profiles as well [48,49]. Gastric adenocarcinoma (GCA) is the most common type of stomach cancer [50]. GCA can be further subdivided into two groups anatomically which are cardia and noncardia adenocarcinomas. Gastric cardia is defined as the tissue that is located at the opening of the esophagus to the stomach and gastric noncardia tissue constitutes the rest of the stomach tissue. These subtypes also exhibit specific expression patterns that differ from each other [51]. Hence, it is of interest to investigate potential differences between the APA patterns of these two distinct subtypes of GCA to better understand the disease etiology and contribute to early detection efforts.

1.5 Aim of the Study

Post-transcriptional regulation of gene expression is affected by numerous factors including APA. APA affects 3' UTR lengths of diverse mRNAs. A trend towards the production of mRNAs with shorter 3' UTRs is observed in diseases that show abnormal rapid proliferation such as cancer. Therefore, we aimed to investigate the APA pattern changes in MDS, glioblastoma and gastric adenocarcinomas which are aggressive proliferative diseases, difficult to diagnose and show poor prognosis, in order to better

explain the pathology of these malignancies. For this purpose, we used microarray experiment datasets containing non-tumor and tumor tissue samples to analyze and compare healthy and disease states.

CHAPTER 2

MATERIALS AND METHODS

2.1 Patient Datasets

Previously performed microarray analysis results for myelodysplastic syndromes, gastric cancer and glioblastoma were obtained from NCBI's Gene Expression Omnibus (GEO). GEO is an archive for high throughput microarray and next-generation sequencing datasets [52]. Patient datasets selected for this study were GSE19429 [53] for MDS, GSE4290 [54], GSE16011 [55], and GSE4271 [56,57] for glioblastoma, and GSE29272 [58] for gastric cancer. GSE12662 [59] and GSE30201 [60] datasets' normal control samples were also included in the analysis.

2.1.1 Myelodysplastic Syndromes - GSE19429, GSE12662 and GSE30201

In GSE19429 dataset, there are 183 MDS patients and 17 healthy control samples collected from various European centers, Duisburg (Germany), Stockholm (Sweden), Oxford and Bournemouth (UK), and Pavia (Italy) [53]. For the study, CD34+ bone marrow cells were isolated from bone marrow samples and used for microarray analysis [53]. Because, MDS is a highly heterogeneous disease in terms of genetic content, only

cytogenetically normal samples were of interest. Cytogenetically abnormal samples were excluded from the analysis and the rest of the samples constituting 94 of total, were sorted into disease subgroups for individual comparisons with normal samples: normal vs. Refractory Anemia (RA), normal vs. Refractory Anemia with Ring Sideroblasts (RARS), normal vs. Refractory Anemia with Excess Blasts 1 (RAEB1) and normal vs. Refractory Anemia with Excess Blasts 2 (RAEB2). Normal CD34+ cell samples were combined with samples from two other datasets which are GSE12662 [59] and GSE30201 [60]. The sample list is given in Appendix Table A.1.

2.1.2 Glioblastoma - GSE4290, GSE16011 and GSE4271

GSE4290 dataset contained a total of 180 samples, 23 of which were non-tumor samples from epilepsy patients and the rest being glioma types ranging from grade II to grade IV [54]. Since, grade IV glioblastomas are the point of interest in this study, 77 GBM samples were separated from the rest of the samples, analyzed and compared to the 23 normal samples. The sample list is given in Appendix Table A.2.

GSE16011 dataset contained a total of 284 samples, 8 of which were normal controls and the rest being glioma samples varying in grades [55]. Again, grade IV GBM samples (n=159) were separated from the other samples, analyzed and compared to the 8 normal samples. The sample list is given in Appendix Table A.3.

GSE4271 dataset contained a total of 100 grade III and IV glioma samples, 23 of which were matched recurrences of 23 tumor samples [56,57]. In order to see whether APA patterns change in tumor recurrence, the 23 samples containing matched recurrences were separated from the rest of the samples, analyzed and compared to their primary tumor samples. The sample list is given in Appendix Table A.4.

2.1.3 Gastric Cancer - GSE29272

This dataset included 134 gastric adenocarcinoma patients admitted to the Shanxi Cancer Hospital between 1998 and 2001. 62 of the patients were diagnosed with cardia and 72 of the patients were diagnosed with noncardia adenocarcinoma. None of them underwent prior therapy. Tumor and matched control tissue samples were obtained by surgical removal of the tumors and surrounding regions. Vital status was followed by re-contact between 2005-2007. Detailed patient information (containing survival data, age, gender, etc.) is given in Appendix Tables A.5 and A.6. [58].

2.2 Alternative Polyadenylation Analysis

In order to find the differences in polyadenylation patterns between healthy and tumor tissues, APADetect was used. APADetect is a probe-based microarray analysis tool devised to identify the genes that are differing in polyadenylation patterns between two sample groups [61]. The tool was developed by Dr. Tolga Can (Department of Computer Engineering, METU, Turkey). APADetect, through available chip designs, analyzes Affymetrix Human Genome U133A Arrays (HGU133A, GPL96) and Affymetrix Human Genome U133 Plus 2.0 Arrays (HGU133Plus2, GPL570). Polyadenylation sites of genes were obtained from PolyA_DB [62], and unique probe sets available on these platforms were divided into two groups as proximal and distal probe sets accordingly. HGU133A contained 2411 probe sets for 2066 genes, whereas HGU133Plus2 contained 3683 probe sets for 3067 genes. Proximal and distal probe intensities were, therefore, indicative of genes' "Short" and "Long" 3' UTR isoforms respectively. From these intensities SLR (Short/Long Ratio) values were calculated and compared between two sample groups, in our case normal vs. disease tissue.

To detect APA events, 4 different filters were applied during analysis in APADetect. First, "size filter" was applied to exclude transcripts that were identified by only one

probe. Second, "outlier probe filter" was applied to exclude individual outlier probes that deviated more than a pre-defined limit from their sample medians. At this stage "size filter" was re-applied. Third, "outlier sample filter" was used. It worked with the same principle as the probe filter to exclude samples that deviate from their group medians. Finally, "distal filter" was applied to exclude the probe sets showing higher distal intensity than proximal intensity. This was done so because the proximal probe set would recognize both short and long isoforms; hence, distal intensity should not be higher than proximal intensity.

2.3 Statistical Analyses

CEL files of the sample datasets were downloaded and analyzed by APADetect by Dr. Tolga Can. APADetect's output is similar to microarray data. However, it only contains SLR values due to alternative polyadenylation event. This output was then analyzed in TIGR MultiExperiment Viewer (MeV) program (http://www.tm4.org/mev.html). MeV is an analysis tool for evaluation of microarray data with several methods such as visualization, classification, clustering, biological motif discovery and statistical analysis [63]. The data obtained from APADetect was log transformed, and then loaded to MeV for statistical analysis. After data was loaded, Percentage Cutoff Filter was applied on the data and genes that did not have expression data in at least 50% of the patients were excluded. In order to find genes showing statistically significant changes in SLR, Significance Analysis for Microarrays (SAM) was applied. SAM performed individual *t* tests for genes in order to detect significant changes in isoform expressions in tumor tissue as compared to normal tissue [64]. Positive and negative significant gene lists were obtained as a result of SAM (Appendix B).

Using SAM, whether APA had any significant relation with different patient characteristics was tested as well. By the detailed patient information table present for gastric cancer patients, whether gender or family history of gastrointestinal tract cancer had any significant effect on altered APA patterns was explored.

Scatter plots for the genes, that were most significantly different with the lowest p-values in terms of APA, were prepared on GraphPad Prism 5 software. Data included log transformed SLR values to avoid dense visualizations of normal samples and disease samples, and to obtain clear plots.

2.4 Survival Analysis

Survival analysis was carried out for the most significantly altered genes in gastric cancer samples. Patients with unknown causes of death were excluded from the analysis; only cancer related deaths were included (41 cardia vs 47 noncardia cancer cases). SLR values of patients were grouped as High (15 patients with highest SLR corresponding to 32% for cardia and 37% of noncardia cases) meaning increased shortening compared to normals and Low (15 patients with lowest SLR) meaning values closer to normals. Statistical significance and p values were determined by log-rank (Mantel-Cox) and Gehan-Breslow-Wilcoxon tests using GraphPad.

2.5 Gene Ontology

Ontology search was performed for the genes with significantly altered APA patterns using DAVID bioinformatics resources [65].

2.6 Cell Lines

MCF7, MDA-MB-231 breast cancer cell lines and AGS gastric cancer cell line were grown for confirmation of the short transcript isoform presence in the cells. All cell lines were grown as monolayers in DMEM growth medium supplemented with 10%

Fetal Bovine Serum and 1% Penicillin/Streptomycin. The incubation conditions were 37^{o}C, 95% humidified air with 5% CO_2.

2.7 Rapid Amplification of cDNA Ends

2.7.1 RNA Isolation

Total RNA was isolated using Isol-RNA Lysis Reagent (5 PRIME). Isolation was performed according to the manufacturer's specifications. 8 ml of Isol-RNA Lysis Reagent was used per one 75 cm^2 tissue culture flask. The reagent was directly applied to the cells. The samples were incubated at room temperature for 5 minutes. After transferring the cells to a sterile, 15-ml tube, 0.2 ml of chloroform is added for 1 ml of Isol Reagent used, meaning a total of 1.6 ml chloroform. The tubes were shaken vigorously for 15 seconds and incubated at room temperature for 2-3 minutes. Then, the aqueous phases containing RNA were obtained by centrifugation at 4500g for 20 minutes at 8^{o}C. The aqueous phases were transferred to new tubes and RNA was precipitated using 4 ml of isopropyl alcohol (0.5 ml isopropanol per 1 ml of Isol Reagent) in each tube. The mixtures were incubated at room temperature for 10 minutes followed by centrifugation at 4500g for 20 minutes at 4^{o}C. The supernatants were removed and the RNA pellets were washed with 8 ml 75% ethanol. After centrifugation at 4500g for 7 minutes at 4^{o}C, the pellets were washed with 70% ethanol, and the final centrifugation was repeated. Finally, the RNA pellets were air-dried for 5-10 minutes. Then, they were rehydrated in 100 µl molecular biology grade water (RNase/DNase-free). Their concentrations were measured using NanoDrop ND1000 (Thermo Scientific).

2.7.2 DNA Contamination Removal and Quantification

For the removal of DNA contamination, total RNA was subjected to DNase I (Roche Applied Science) treatment for 1 hour at 37^{o}C.Then, phenol-chloroform extraction was applied. Reaction conditions are given in Table 2.1.

Table 2.1 DNase I Treatment Reaction Mixture

Reagents	**Amount**
Total RNA	5 µg
10X Incubation Buffer (400 mM Tris-HCl, 100 mM NaCl, 60 mM $MgCl_2$, 10 mM $CaCl_2$, pH 7.9)	10 µl
DNase I Recombinant, RNase-free (10 units/µl) Roche Applied Sciences (Catalogue No: 04 716 728 001)	2 µl
Molecular Biology Grade Water (RNase-free)	To 100 µl

RNA samples were quantified spectrophotometrically using NanoDrop after phenol-chloroform extraction (Appendix C). Lack of DNA contamination was confirmed by Glyceraldehyde 3-Phosphate Dehydrogenase house-keeping gene specific primers' usage in PCR. The primer sequences were as follows:

GAPDH_F: 5'-GGGAGCCAAAAGGGTCATCA-3'

GAPDH_R: 5'-TTTCTAGACGGCAGGTCAGGT-3'

Expected product size is 409 bp. PCR conditions were: initial denaturation at 94 ^{o}C for 2 minutes was followed by 40 cycles of 94 ^{o}C-30 seconds denaturation, 56 ^{o}C-30 seconds annealing and 72 ^{o}C-30 seconds extension. Final extension was at 72 ^{o}C for 10 minutes. Genomic DNA was used as positive control.

2.7.3 3' RACE, Cloning and Sequencing

RACE specific cDNAs were synthesized in 20 µl volume using the 3'-RACE Kit(Roche) with oligo dT-anchor primer (5'-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTV-3'). For PCR forward gene specific and reverse anchor specific primers were used. 3 µg of MDA-MB-231 RNA, 2 µg of MCF7 RNA and 1.5 µg of AGS RNA were used in cDNA synthesis.

TCF3 short isoform's 3'-RACE product, whose expected size was 450 bp, was PCR amplified using gene specific forward primer and anchor specific reverse primer on cell line RACE specific cDNAs. The primer sequences and PCR conditions were as follows:

TCF3-F-5'- CAAAACCTGAAAGCAAGCAA -3'

TCF3-R-5'- TTAGGCACAATTTGCTGGTG -3'

Anchor-R-5'- GACCACGCGTATCGATGTCGAC -3'

Table 2.2 3'-RACE PCR Conditions for *TCF3*

	Initial Denaturation (2 minutes)	Denaturation (30 seconds)	Annealing (30 seconds)	Extension (30 seconds)	Cycle	Final Extension (10 minutes)
TCF3	94^{o}C	94^{o}C	62^{o}C	72^{o}C	35	72^{o}C

PCR amplified products were loaded on a 2% Agarose gel, run at 100V for 30-60 minutes and visualized by UV exposure. Bands of confirmed sizes from AGS cell line cDNA were cut from the gel and DNA was extracted using Zymoclean Gel DNA Recovery Kit (Zymo Research) according to manufacturers' protocol. The fragments were then quantified using NanoDrop.

Purified PCR products were ligated into pGEM-T Easy Vector System (Promega). Insert:vector ratio was 1:1. XL1-Blue *E. coli* competent cells were a kind gift from Muyan Lab. Competent cells were transformed with 2 µl and 5 µl ligation products separately. Bacteria and ligation product mixtures were incubated on ice for 1 hour, followed by 45 second heat shock at 42^{o}C and a further 5 minute incubation on ice. Bacteria were grown in 500 µl SOC medium for 1 hour at 37^{o}C with shaking at 200 rpm. Then, they were inoculated on agar plates containing ampicillin, X-gal and IPTG for both antibiotic and blue/white selection. Colonies that appeared white were presumed to have undergone transformation with vector containing insert.

Colony PCR was used to confirm presence of insert in positive colonies. Forward and reverse primers used to amplify the short isoform were used in colony PCR to check for the availability insert. Purified PCR products used in cloning were used as positive control for size confirmation. After colony screening with PCR, 2 colonies for TCF3 were chosen to grow in LB Broth growth medium. Colonies were grown overnight at 37^{o}C in a shaker incubator with a speed of 200 rpm. Plasmid isolation was performed

using High Pure Plasmid Isolation Kit (Roche). After quantification with NanoDrop, 500 ng plasmid was used in double restriction enzyme digestion to confirm insert presence and correct size. Apa I and Sac I restriction enzymes were used for double digestion in Buffer B, at 37^{o}C for 2 hours. Finally, the plasmids were sent to sequencing.

CHAPTER 3

RESULTS AND DISCUSSION

The aim of this study was to analyze different cancer types for altered APA patterns. For this purpose, NCBI's GEO database, (http://www.ncbi.nlm.nih.gov/geo/) [52] was used to gather patient sample datasets that can be compared to normal control samples in terms of their poly(A) site usage. Myelodysplastic syndromes, glioblastoma and gastric cancers were the main focus of this study, because all are aggressive and/or proliferative diseases that lead to low overall survival and early diagnosis is difficult. GSE19429 [53], GSE12662 [59] and GSE30201 [60] series were used for MDS; GSE4290 [54], GSE16011 [55], and GSE4271 [56,57] series were used for glioblastoma, and GSE29272 [58] series were used for gastric cancer.

3.1 APA in Myelodysplastic Syndromes (MDS)

Given that APA patterns are altered significantly during neoplastic transformation [5] and MDS is a group of bone marrow diseases that is associated with a risk of transformation into acute myeloid leukemia [38], we wanted to investigate any potential difference between MDS subtypes that could possibly aid easier diagnosis, subtyping or prognosis prediction of this disease. MDS is a highly heterogenous disease in terms of genetics [40]. Although certain structural chromosomal abnormalities are commonly seen in MDS and

leukemia [41], cytogenetically normal cases of MDS can also progress into leukemia [40]. Therefore, it was of interest to investigate cytogenetically normal MDS patients in order to see whether APA could play a potential role in AML conversion. Cytogenetically normal MDS samples were divided into disease subgroups that they progressed into. The expression data for cytogenetically normal MDS samples of GSE19429 dataset [53] was re-analyzed using the probe-based APADetect tool in accordance with the poly(A) positions present in the genome (polyA_DB) [66]. The chosen patient samples were re-analyzed according to proximal and distal probes grouped based on the poly(A) positions. Probe set intensity means of the patient groups were compared to those of the normal sample group. Normal samples were CD^{34+} hematopoietic stem cells isolated from the bone marrows of healthy individuals, and these were compared to CD^{34+} hematopoietic stem cells isolated from patient bone marrow samples. Alterations in the proximal to distal probe ratio were used as indicators of changes in short and long 3' UTR levels. To explore the statistical significance of SLR (Short/Long Ratio) values of patients as compared to normal (healthy) samples, SAM (Significance Analysis of Microarrays) [64] was used. The MDS subgroups analyzed were refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), refractory anemia with excess blasts-1 (RAEB1), and refractory anemia with excess blasts-2 (RAEB2).

3.1.1 APA in Refractory Anemia (RA)

RA is the subtype of MDS that is described as drug-resistant anemia. It is characterized by a shortage of one type of blood cell lineage, and a slight increase in the number of immature blood cells in blood and bone marrow [67]. RA group of patients rarely progress into AML [52].

Cytogenetically normal RA patient samples were retrieved from GSE19429 dataset [53] for APA analysis in comparison to normal samples. This dataset contained 22 RA patient samples and 17 normal samples. In order to increase the number of normal samples, GSE12662 dataset's 5 normal samples [59], and GSE30201 dataset's 11 normal

samples were also included in the analysis [60]. SAM was used to investigate significant shortening and lengthening events in RA patients (n=22) as compared to normal samples (n=33) (Figure 3.1).

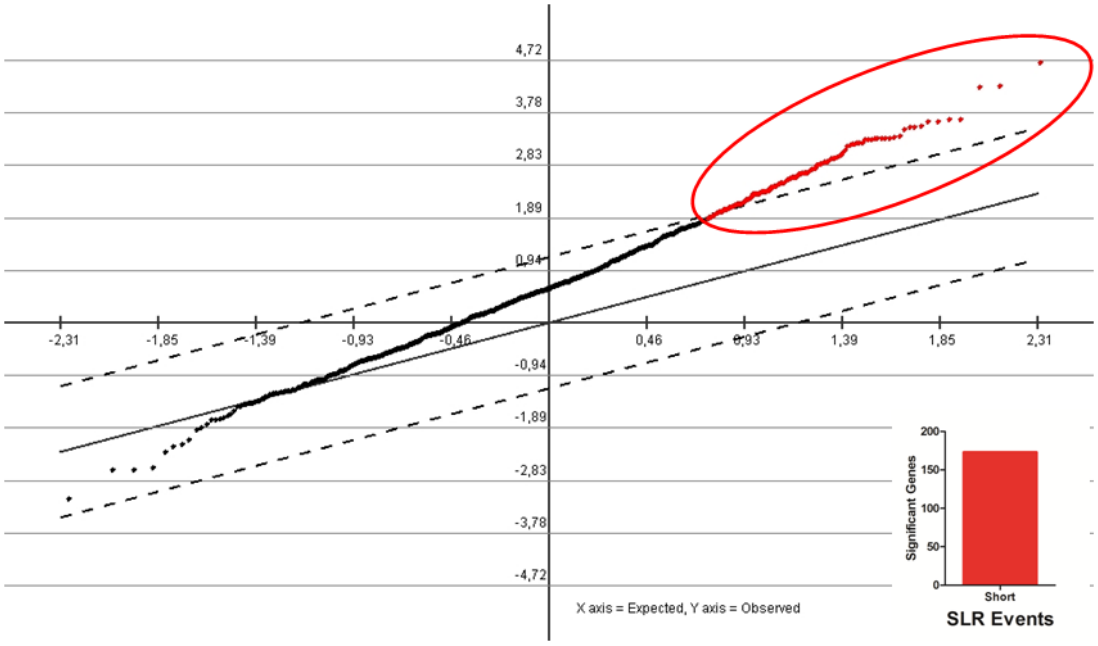


Figure 3.1 SAM graph of normal samples (n=33) vs. RA (n=22). Each red dot indicates a significantly higher SLR gene transcript in patient samples as opposed to normal samples. Higher SLR is indicative of shortened transcripts. In this case, there were 173 significant shortening events. Genes were identified as statistically significant if q-values were <2%. The q-value is an adaptation of p-value for large scale statistical analysis such as SAM. It refers to "False Discovery Rate (FDR)" of genes, so it is a measure of the risk of incorrectly identifying a gene as significant [68].

SAM results indicate 173 shortened transcripts (Appendix Table B.1). Scatter plots were prepared for the 6 most significantly shortened genes in the patients (*SCARB2, RTN1, CFLAR, APIP, SNX11 and FLII*) using individual data obtained from APADetect (Figure 3.2).

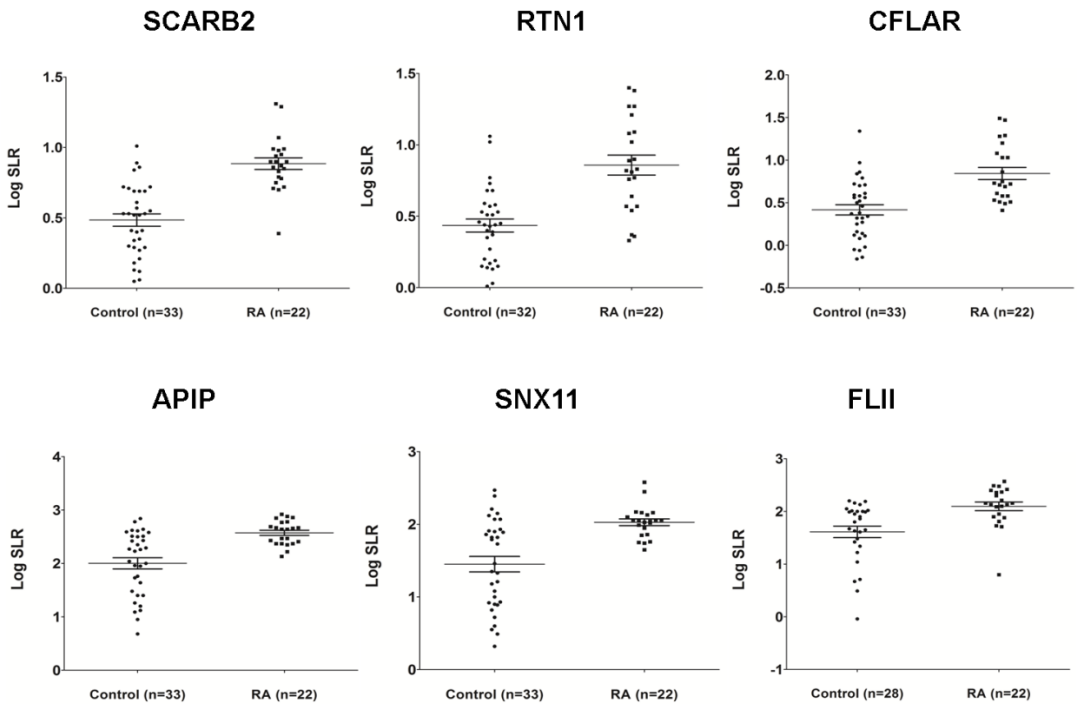


Figure 3.2 Scatter plots of 6 genes that undergo significant 3'UTR shortening in RA (n=22) patients as compared to normal samples (n=33). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

It can be seen from the scatter plots (Figure 3.2) that the genes that have significant changes in SLR values at disease states according to SAM, really do deviate largely from healthy controls in comparison. Increased SLR values indicate either an abundance of short transcripts, or a shortage of long transcripts. Table 3.1 gives detailed information about these genes' functions.

Table 3.1 Genes undergoing 3' UTR shortening in RA

Gene	Function/Relevance to Cancer	References
SCARB2 (Scavenger receptor class B, member 2)	-glycoprotein -present in endosomal/lysosomal membranes -membrane transportation	69
RTN1 (Reticulon 1)	-reticulon gene family -endoplasmic reticulum localized -neurological disease/cancer marker	70
CFLAR (CASP8 and FADD-like apoptosis regulator)	-apoptosis regulator	71
APIP (APAF1 interacting protein)	-negative regulator of hypoxic injury	72
SNX11 (Sorting nexin 11)	-sorting nexin protein family -intracellular trafficking	73
FLII (Flightless I homolog)	-gelsolin-like actin binding protein -high levels in bone marrow -ability to bind nuclear receptors	74

As illustrated in Table 3.1, significantly shortened genes' functions are important in critical processes of the cell such as metabolism, cell cycle, apoptosis as well as specific bone marrow activity. This is important, because it goes to show how APA may be contributing to the progression of RA subgroup of MDS.

It is also of interest to note that all SLR changes were in the direction of shortening; therefore, suggesting potential oncogenic roles for these genes.

3.1.2 APA in Refractory Anemia with Ringed Sideroblasts (RARS)

RARS is also a drug-resistant subtype of MDS. In addition to showing the same symptoms as the RA subtype, the immature cells seen in RARS cannot incorporate iron correctly in hemoglobin leading to formation of unusual ferritin structures in the

cytoplasm [67]. RARS group of patients show similar prognosis to RA group of patients and about 1-2% of patients progress to AML [52].

Cytogenetically normal RARS patient samples were retrieved from GSE19429 dataset [53] for APA analysis in comparison to normal samples. This dataset contained 31 RARS patient samples. The aforementioned normal samples (33 in total) are included in the analysis for comparison. SAM was used to investigate significant shortening and lengthening events in RARS patients as compared to normal samples. (Figure 3.3).

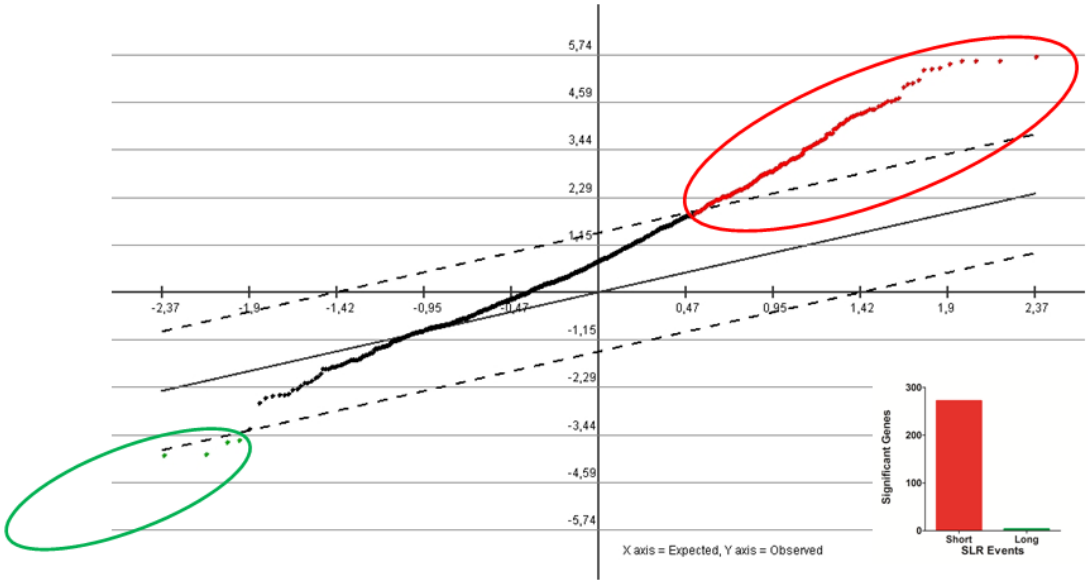


Figure 3.3 SAM graph of normal samples (n=33) vs. RARS (n=31). In the comparison between normal samples and RARS, the red and green dots indicate shortened and lengthened gene transcripts, respectively. In this case, there were 272 shortening and 4 lengthening events. All significant genes were identified as statistically significant with q-values<2%.

SAM results indicate 272 shortened and 4 lengthened transcripts (Appendix Tables B.2 and B.3). Scatter plots were prepared for the 6 most significantly shortened (*SLC48A1, GLRX3, GTPBP8, YARS2, SF4 and C4A*) and 4 lengthened genes (*MTHFD2, PABPC3, MRP63 and STMN1*) using individual data obtained from APADetect (Figures 3.4 and 3.5).

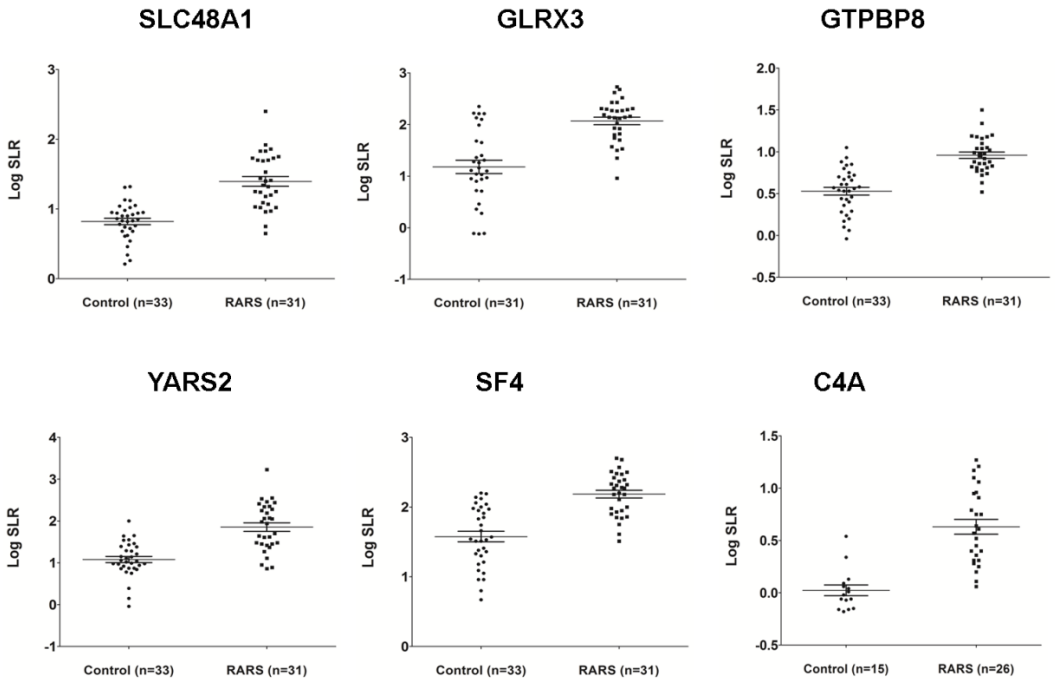


Figure 3.4 Scatter plots of 6 genes that undergo significant 3'UTR shortening in RARS patients (n=31) as compared to normal samples (n=33). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

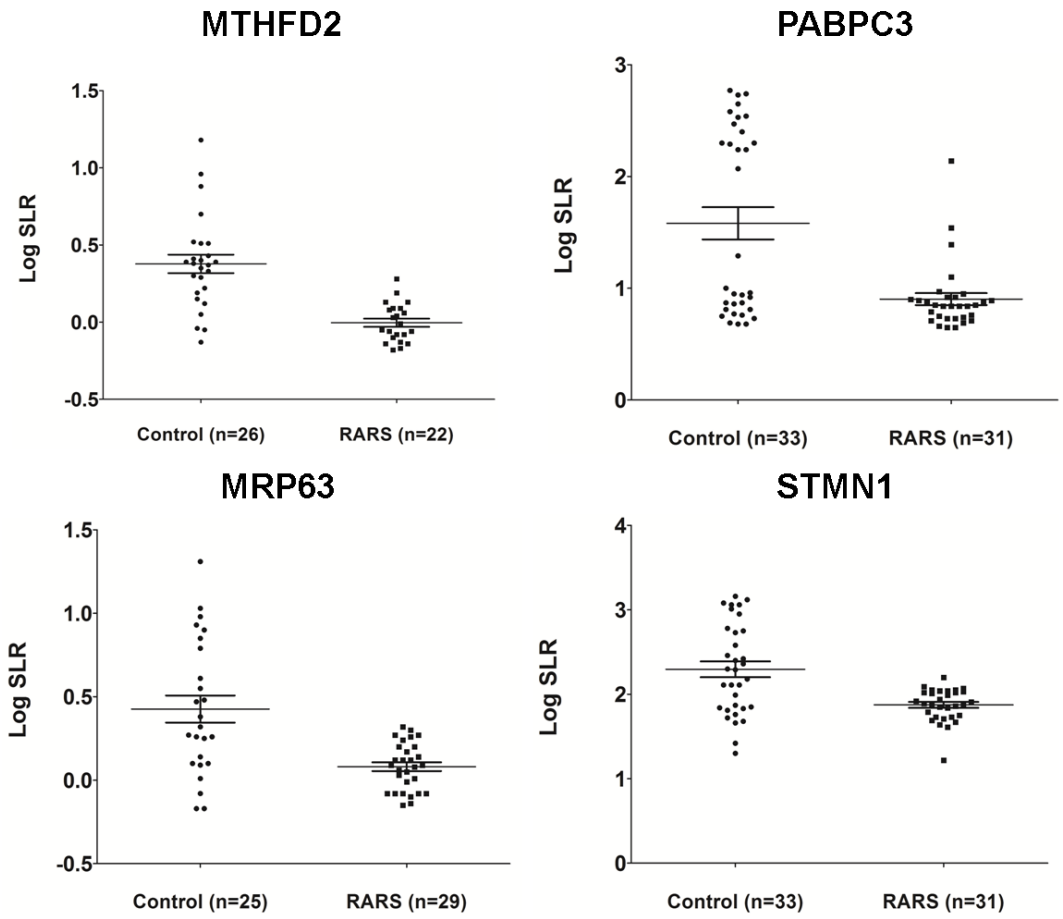


Figure 3.5 Scatter plots of 4 genes that undergo significant 3'UTR lengthening in RARS patients (n=31) as compared to normal samples (n=33). The scatter plots illustrate the 4 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

The scatter plots in Figures 3.4 and 3.5 clearly show that the genes with significantly different SLR values according to SAM, really do deviate largely from healthy controls in comparison. As opposed to increased SLR values, decreases in SLR indicate either an abundance of long transcripts, or a shortage of short transcripts. Table 3.2 gives detailed information about these genes' functions.

Table 3.2 Genes undergoing 3' UTR shortening/lengthening in RARS

Gene	Function/Relevance to Cancer	References
Shortening Events		
SLC48A1 (Solute carrier family 48, member 1)	-membrane-bound heme transporter -suppression causes impaired endocytosis	75 76
GLRX3 (Glutaredoxin 3)	-oxidoreductase enzyme -potential apoptosis inhibitor -potential cancer marker	77
GTPBP8 (GTP-binding protein 8 (putative))	-undefined	
YARS2 (Tyrosyl-tRNA synthetase 2, mitochondrial)	-mitochondrial enzyme for tyrosine-tRNA attachment -associated with sideroblastic anemia	78
SF4 (SURP and G patch domain containing 1)	-SURP domain containing -splicing factor	79
C4A (Complement component 4A)	-complement system protein -immune system activity	80
Lengthening Events		
MTHFD2 (methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase)	-bifunctional mitochondrial enzyme -elevated expression in cancer -poor prognosis in breast cancer	81 82
PABPC3 (Poly(A) binding protein, cytoplasmic 3)	-mRNA stability, translation initiation regulator	83
MRP63 (Mitochondrial ribosomal protein 63)	-mitochondrial ribosome component	84
STMN1 (Stathmin 1)	-cytosolic phosphoprotein -microtubule destabilizer -related with high-risk MDS	85

Table 3.2 also shows significantly shortened genes' functions to be important in cell metabolism, cell division and specific bone marrow activity. This supports the suggestion that there may be a link between APA and MDS.

3.1.3 APA in Refractory Anemia with Excess Blasts-1 (RAEB1)

RAEB1 is the subtype of MDS that shows, in addition to drug-resistant anemia, at least 5 fold increase in immature blood cells [67]. It is characterized by a shortage of 2 or more types of blood cell lineages, and a high increase in the number of immature blood cells in blood and bone marrow [67]. About 25% of patients show progression to AML [52].

Cytogenetically normal RAEB1 patient samples were retrieved from GSE19429 dataset [53] for APA analysis in comparison to normal samples. This dataset contained 21 RAEB1 patient samples. The aforementioned normal samples (33 in total) are included in the analysis for comparison. SAM was used to investigate significant shortening and lengthening events in RAEB1 patients as compared to normal samples. (Figure 3.6).

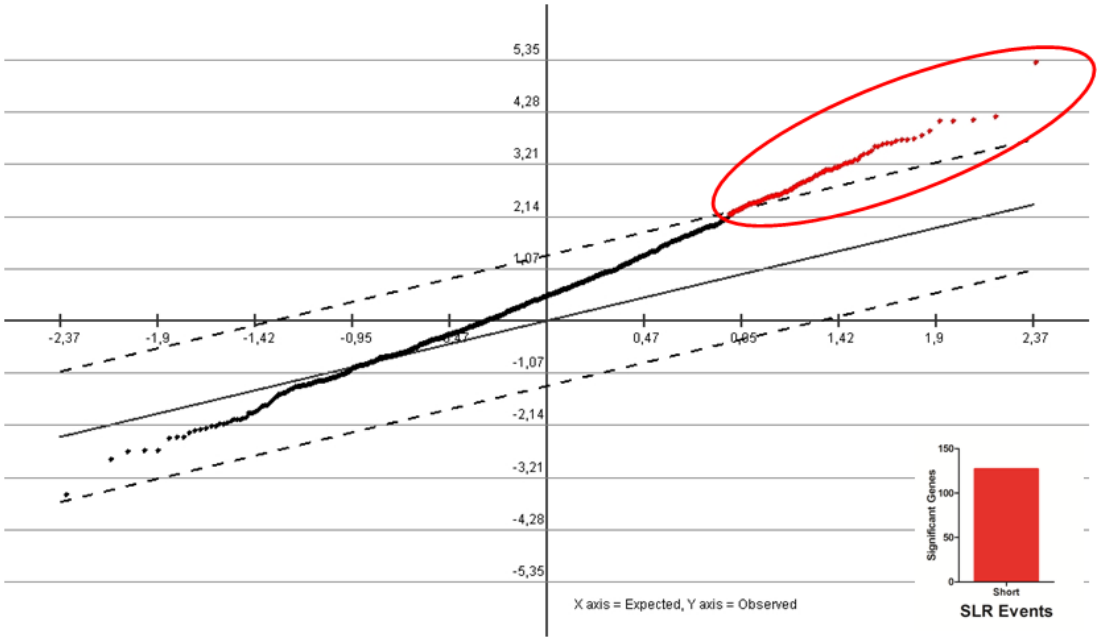


Figure 3.6 SAM graph of normal samples (n=33) vs. RAEB1 (n=21). Each red dot indicates a shortened gene transcript in patient samples as opposed to normal samples. In this case, there were 127 shortening events. All significant genes were identified as statistically significant with q-values<2%.

SAM results indicate 127 shortened transcripts (Appendix Table B.4). Scatter plots were prepared for the 6 most significantly altered genes (*SLC46A3, ZNF554, ZAK, ZNF625, SPOCK3 and C9orf85*) using individual data obtained from APADetect (Figure 3.7).

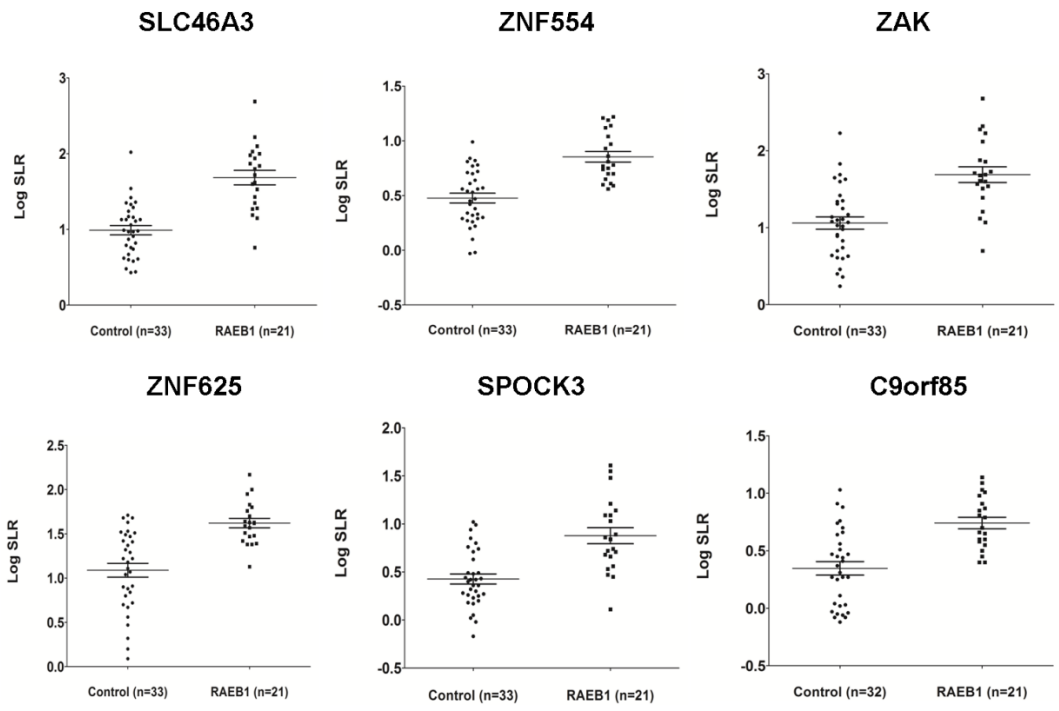


Figure 3.7 Scatter plots of 6 genes that undergo significant 3'UTR shortening in RAEB1 patients (n=21) as compared to normal samples (n=33). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

It can be seen from the scatter plots (Figure 3.7) that the genes that have significant changes in SLR values at disease states according to SAM, really do deviate largely from healthy controls in comparison. Table 3.3 gives detailed information about these genes' functions.

Table 3.3 Genes undergoing 3' UTR shortening in RAEB1

Gene	Function/Relevance to Cancer	References
SLC46A3 (Solute carrier family 46, member 3)	-undefined	
ZNF554 (Zinc finger protein 554)	-undefined	
ZAK (Sterile alpha motif and leucine zipper containing kinase AZK)	-MAPKKK family protein -cell cycle checkpoint regulator -pro-apoptotic activity	86
ZNF625 (Zinc finger protein 625)	-undefined	
SPOCK3 (sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3)	-proteoglycan -related to T-cell leukemia	87
C9orf85 (Chromosome 9 open reading frame 85)	-undefined	

Although most of the genes in Table 3.3 do not have defined functions, still bone marrow and cell cycle related genes can be found in SAM results suggesting an increased likelihood of finding a link between APA and MDS.

3.1.4 APA in Refractory Anemia with Excess Blasts-2 (RAEB2)

RAEB2 subtype of MDS shows the same symptoms as RAEB1 with even more increased immature blood cells [67]. In addition, formation of Auer rods can be observed [67]. Auer rods are abnormal structures derived from material that can readily be found in the cytoplasmic granules and are commonly seen in malignant blood diseases [88]. Approximately 33% of patients progress into AML [52].

Cytogenetically normal RAEB2 patient samples were retrieved from GSE19429 dataset [53] for APA analysis in comparison to normal samples. This dataset contained 20 RAEB2 patient samples. The aforementioned normal samples (33 in total) are included in the analysis for comparison. SAM was used to investigate significant shortening and lengthening events in RAEB2 patients as compared to normal samples. (Figure 3.8).

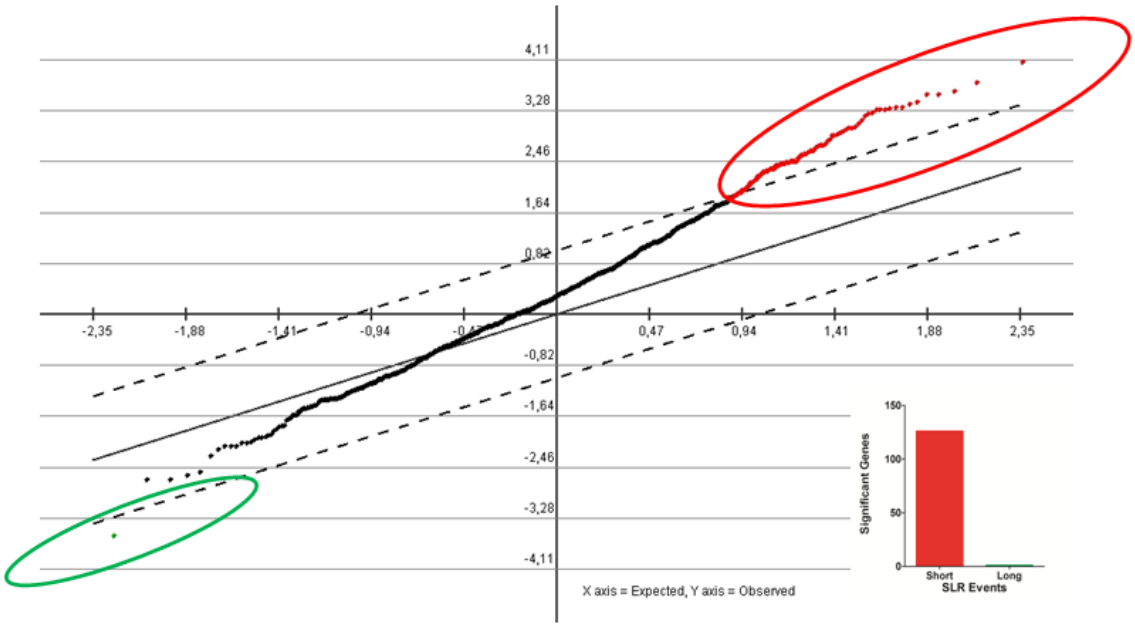


Figure 3.8 SAM graph of normal samples (n=33) vs. RAEB2 (n=20). In the comparison between normal samples and RAEB2, the red and green dots indicate shortened and lengthened gene transcripts, respectively. In this case, there were 126 shortening and 1 lengthening events. All significant genes were identified as statistically significant with q-values<2%.

SAM results indicate 126 shortened and 1 lengthened transcripts (Appendix Tables B.5 and B.6). Scatter plots were prepared for the 6 most significantly shortened (*SLC46A3, ADRBK1, SF4, STAC, LGI4 and TET2*) and 1 lengthened gene (*PABPC3*) using individual data obtained from APADetect (Figures 3.9 and 3.10).

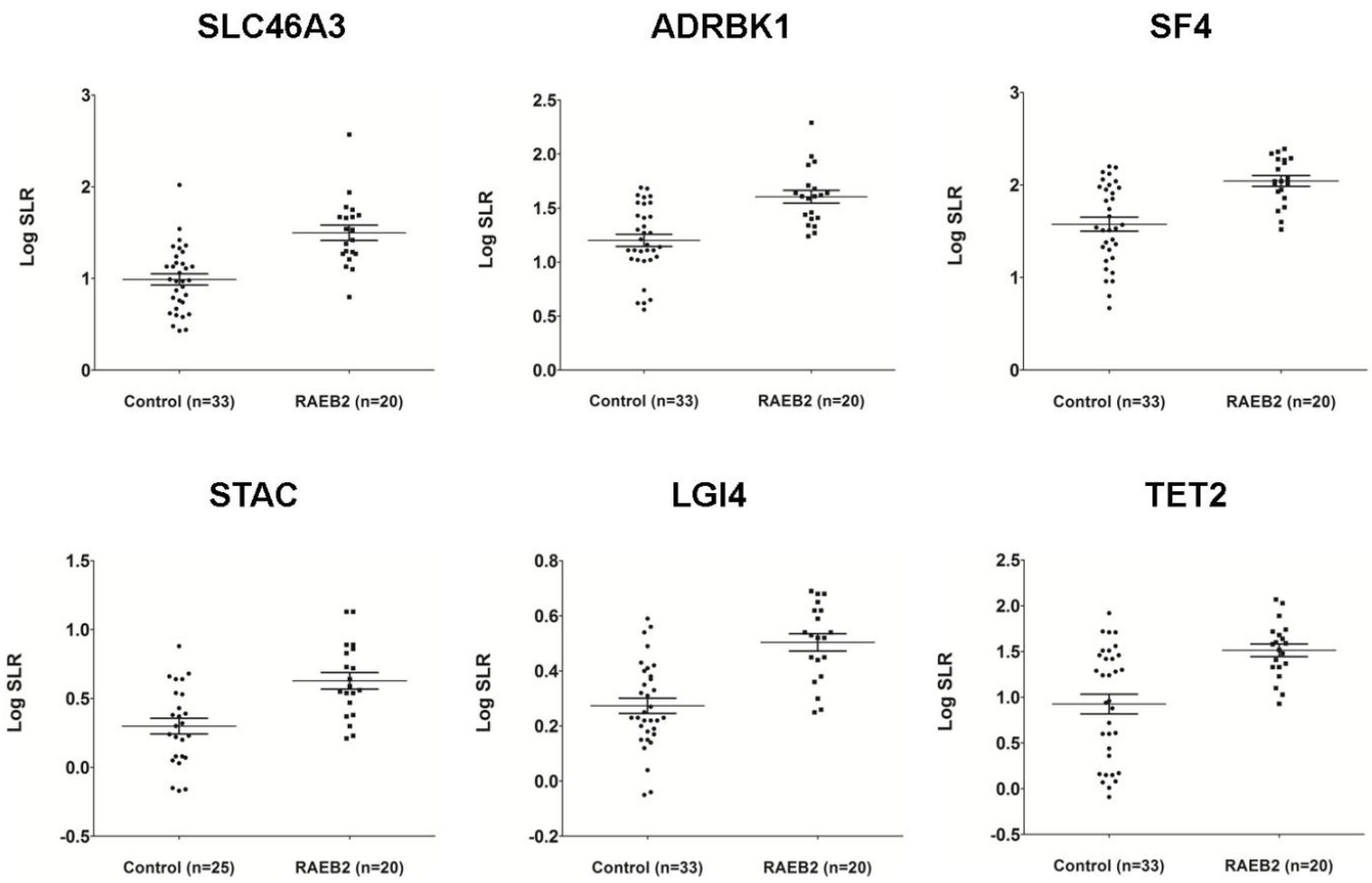


Figure 3.9 Scatter plots of 6 genes that undergo significant 3'UTR shortening in RAEB2 patients (n=20) as compared to normal samples (n=33). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

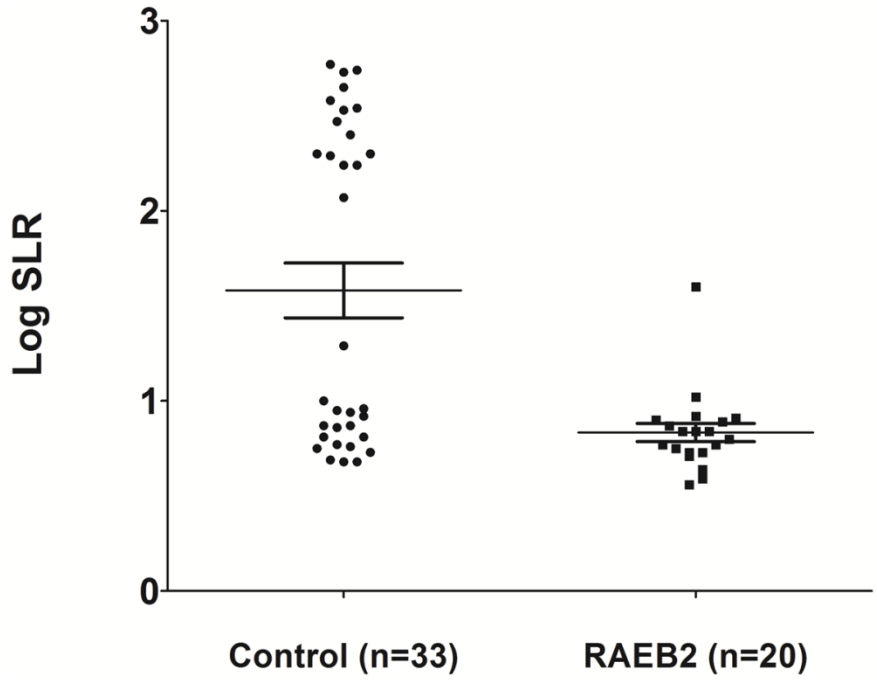


Figure 3.10 Scatter plot of *PABPC3* gene that undergoes significant 3'UTR lengthening in RAEB2 patients (n=20) as compared to normal samples (n=33). The scatter plot illustrates the *PABPC3* gene that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

The scatter plots in Figures 3.9 and 3.10 clearly show that the genes with significantly different SLR values according to SAM, really do deviate largely from healthy controls in comparison. Table 3.4 gives detailed information about these genes' functions.

Table 3.4 Genes undergoing 3' UTR shortening/lengthening in RAEB2

Gene	Function/Relevance to Cancer	References
Shortening Events		
SLC46A3 (Solute carrier family 46, member 3)	-undefined	
ADRBK1 (Adrenergic, beta, receptor kinase 1)	-desensitizer of G-protein coupled receptors -related to cell cycle, survival, migration	89
SF4 (SURP and G patch domain containing 1)	-SURP domain containing -splicing factor	79
STAC (SH3 and cysteine rich domain)	-putative adaptor -neuronal expression	90
LGI4 (leucine-rich repeat LGI family, member 4)	-neuronal expression -possible link to epilepsy	91
TET2 (Tet methylcytosine dioxygenase 2)	-myelopoiesis -irregularities associated with myeloproliferative diseases	92 93
Lengthening Event		
PABPC3 (Poly(A) binding protein, cytoplasmic 3)	-mRNA stability, translation initiation regulator	83

The profile seen in Table 3.4 fits well with the findings so far which suggest a potential role for APA in MDS disease progression. It is important to note that genes identified as significantly altered in terms of APA are turning out to be related to both tissue specific functions, and cell cycle phases.

Interestingly in MDS, the alterations in APA patterns seem to be tending towards a general shortening of transcripts with few lengthening events. Such a tendency to lack lengthening events was not seen in APA analyses of solid tumors such as glioblastoma or gastric cancer (discussed later). This could be important in that a tendency towards shortening appears to be specific to proliferative or malignant blood diseases. Therefore, exploration of such patterns in other blood malignancies such as different leukemia types could provide valuable information about the

biological events underlying these diseases or even provide potential biomarkers for diagnosis or targeted therapy.

It is also important to note that when the genes that show significantly altered APA patterns in different subtypes of MDS are compared, many of them appear to be commonly altered. Table 3.5 is the list of these commonly altered genes. As there were few significant lengthening events in MDS, all commonly altered genes are undergoing significant shortening.

Table 3.5 Genes commonly undergoing 3' UTR shortening in MDS subtypes. The table contains common 3' UTR shortening events out of 173 shortening events in RA, 272 shortening events in RARS, 127 shortening events in RAEB1 and 126 shortening events in RAEB2.

3'UTR shortening					
YARS2	*SCARB2*	*RTN1*	*FLII*	*POLR2C*	*C9orf82*
GTPBP8	*ACTL6A*	*GLRX3*	*SLC46A3*	*ZKSCAN5*	*SNX13*
SPOCK3	*NNT*	*NIP30*	*LYPLA2*	*BCLAF1*	*CREB1*
SF4	*C9orf85*	*ADRBK1*	*EBNA1BP2*	*VPS4B*	*ATP7A*
KLHL20	*KLHL24*	*ENOX2*	*ZAK*	*RB1*	*TDGF1*
MOBP	*HACE1*	*CTBS*	*ITGA11*	*SNHG10*	*ZNF554*
ZNF625	*CREBBP*	*NIPBL*	*C10orf119*	*LAS1L*	*UNKNOWN* (235386_at)
BCLAF1 (different probeset)					

MDSs originate from the inability of $CD34^+$ hematopoietic stem cells (HSCs) to differentiate into mature blood cell lineages. Therefore, deficiency in multiple blood cell lineages is observed [94]. The extent of the shortening events that we observed may have been a result of the increased proliferation rates occuring to compensate for the lack of blood cells. Considering that HSCs are actively proliferating under normal circumstances [95], further increase in their proliferation rates may be causing the lack of lengthening events in MDS subtypes.

The presence of numerous commonly altered genes in the examined MDS subtypes, regardless of AML transformation risk, indicates that changes in APA events start occurring as early events during the progression of MDS. This is highly important, because identification of such early changes in certain genes can be a means of early diagnosis in future prospect. In addition, certain genes showing differential APA event changes between subtypes, may be used in subtyping of the syndromes at the molecular level.

In addition to APADetect and SAM analyses, ontology analysis was also performed for all genes with significantly altered APA patterns in MDS to better evaluate a possible role of APA (Figure 3.11).

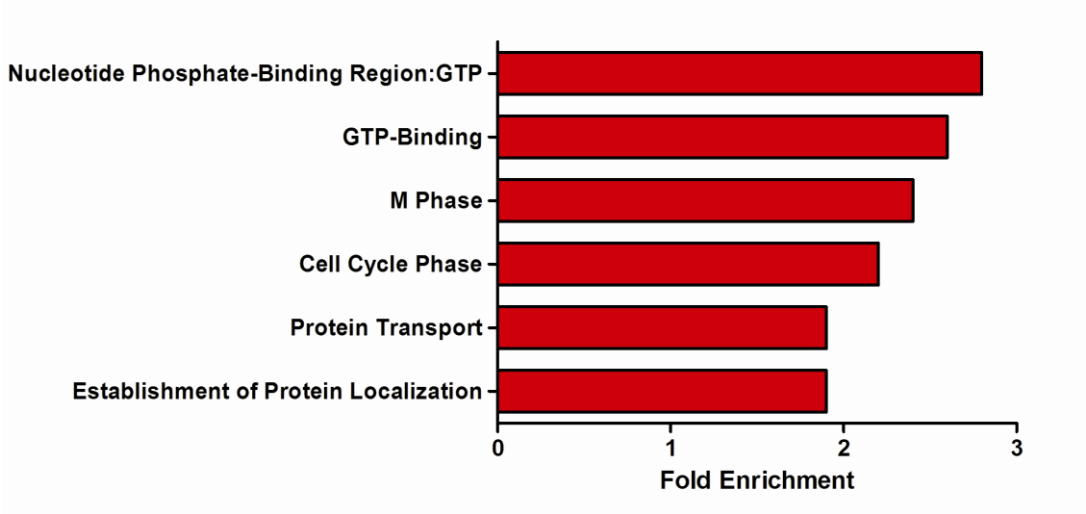


Figure 3.11 Enrichment graph of 297 genes showing significantly altered APA in MDS subtypes in terms of both shortening and lengthening. Gene ontology search was carried out for all the significantly altered genes in MDS in terms of APA. DAVID results show enrichment in genes having functions related to signaling or cell cycle stages. The gene clusters were obtained using Functional Annotation Clustering Tool. The most significant cluster had an enrichment score of 2.44, containing 13-15 GTP-binding protein genes. The second most significant cluster had an enrichment score of 2.32, containing 25-26 protein localization genes. The third most significant cluster had an enrichment score of 2.24, containing 14-19 cell cycle protein genes. Enrichment scores were obtained using the high stringency filter of the software. (p-value<0.001 for each cluster)

As can be seen from Figure 3.11, significantly altered transcripts (mostly shortened) are enriched in that an increase of around 2-fold is seen in genes that have important functions in cell cycle stages and cell division. Since enrichment scores above 1.3 is considered to be significant [96], and the p-values for each enriched cluster were less than 0.001, the enrichment patterns were significant. This goes to show how important APA pattern changes can be in proliferative and cancerous diseases. These results may be more significant if the number of analyzed genes could be increased by studying larger patient datasets. Also, functional significance of the enrichments needs to be verified further.

In order to check the validity of our approach, scatter plots were prepared for genes whose SLR values were equal to one. As expected, no significant difference between control samples and patient samples could be observed (Figure 3.12).

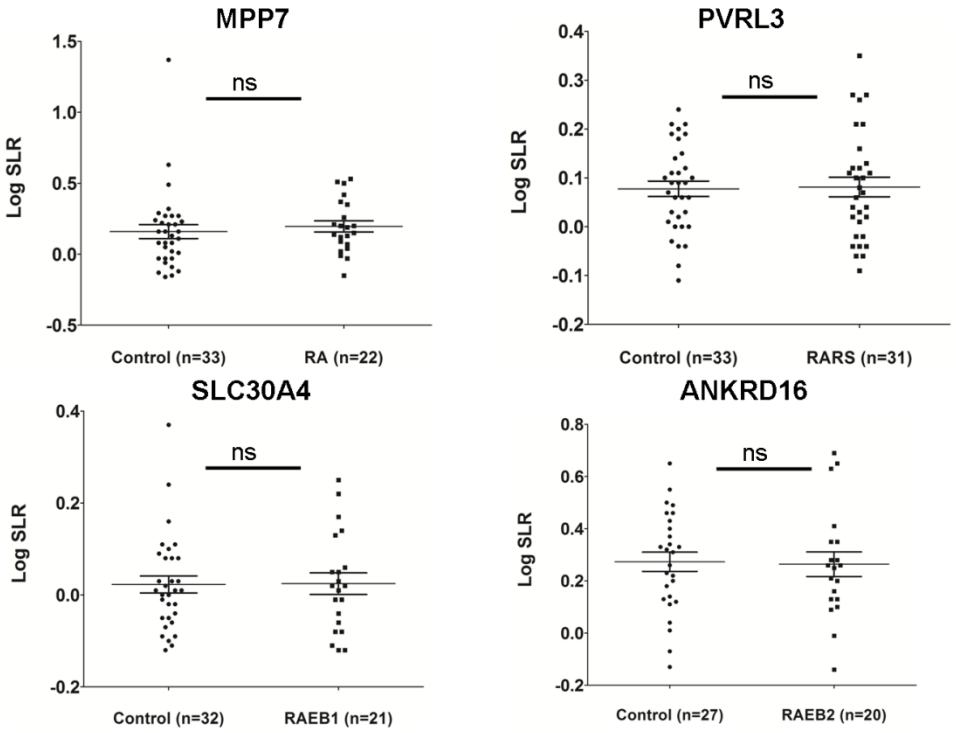


Figure 3.12 Scatter plots of 4 genes that do not undergo 3'UTR alteration in MDS subtypes as compared to normal samples. The scatter plots illustrate 4 genes that showed no significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

3.2 APA in Glioblastoma (GBM)

APA changes in glioblastoma were also of interest, because GBM is the most common and lethal type of brain tumors [43]. Despite different therapy approaches (surgical resection, radiation, chemotherapy), GBM remains incurable due to aggressiveness and invasiveness leading to recurrence [43]. The datasets used for APA analysis were GSE4290 [54], GSE16011 [55] and GSE4271 [56]. We preferred to analyze the datasets separately, because when we compared the control samples of the datasets to each other we saw significant differences in between them. This may have been resulting from the fact that the control samples of GSE4290 [54] were obtained from epilepsy patients. Although the samples were non-tumor, there may have been epilepsy specific APA events, leading to the differences that we observed. Also the control samples may have been acquired from different regions of the brain which may manifest alternating APA patterns due to anatomical site differences.

GSE4290 dataset contained 23 non-tumor samples from epilepsy patients with no tumors, and 77 glioblastoma tumor samples [54]. GSE16011 dataset contained 8 normal samples from normal adult brain, and 159 glioblastoma tumor samples [55]. GSE4271 dataset contained 100 samples, 23 of which were given together with their matched recurrences [57]. Those 23 samples were analyzed and compared to their matched recurrences.

3.2.1 APA in GSE4290 Dataset

This dataset contained a total of 180 samples, 23 of which were non-tumor samples from epilepsy patients. 77 of the glioma samples were primary glioblastoma tumor samples. Therefore, the analysis and comparison were carried out using the 23 control and 77 tumor samples.

After APADetect analysis, firstly, tumor tissue samples were compared to control samples in SAM (Figure 3.13).

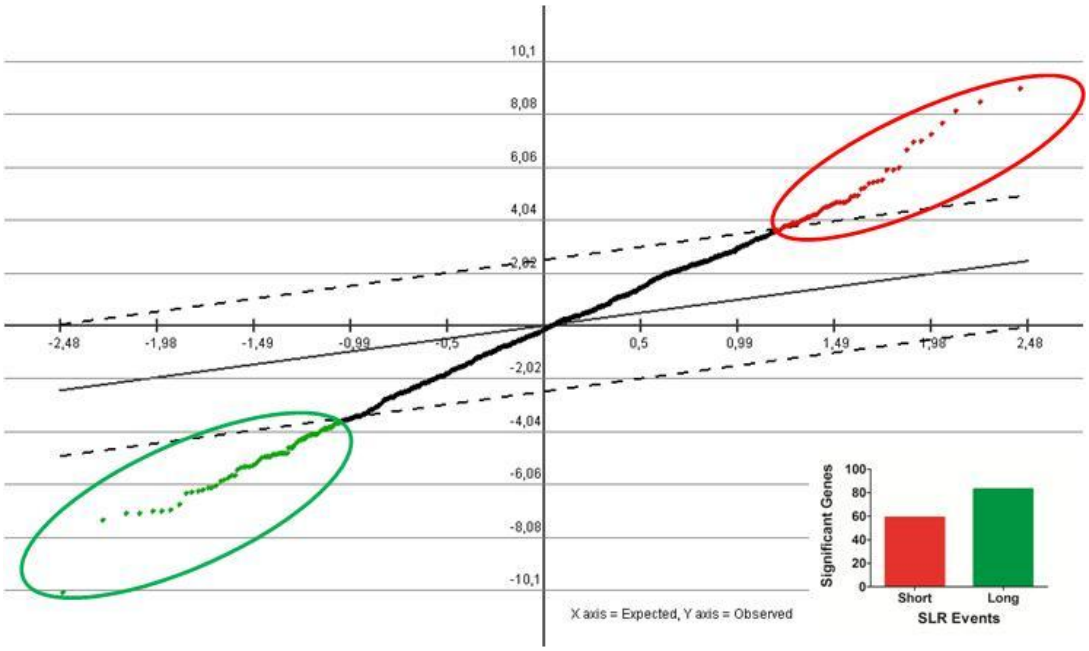


Figure 3.13 SAM graph of normal samples (n=23) vs. Glioblastoma tumor samples (n=77) in GSE4290. In the comparison between normal samples and tumor, the red and green dots indicate shortened and lengthened gene transcripts, respectively. In this case, there were 59 shortening and 83 lengthening events. All significant genes were identified as statistically significant with q-values<2%.

SAM results indicate 59 shortened and 83 lengthened transcripts (Appendix Tables B.7 and B.8). Scatter plots were prepared for the 6 most significantly shortened (*KIAA1245, SNX3, RPL13, APP, ESF1 and KRIT1*) and 6 lengthened genes (*KPNA1*,

DLGAP5, LOC157627, SSX2IP, PSMC6 and EIF5B) using individual data obtained from APADetect (Figures 3.14 and 3.15).

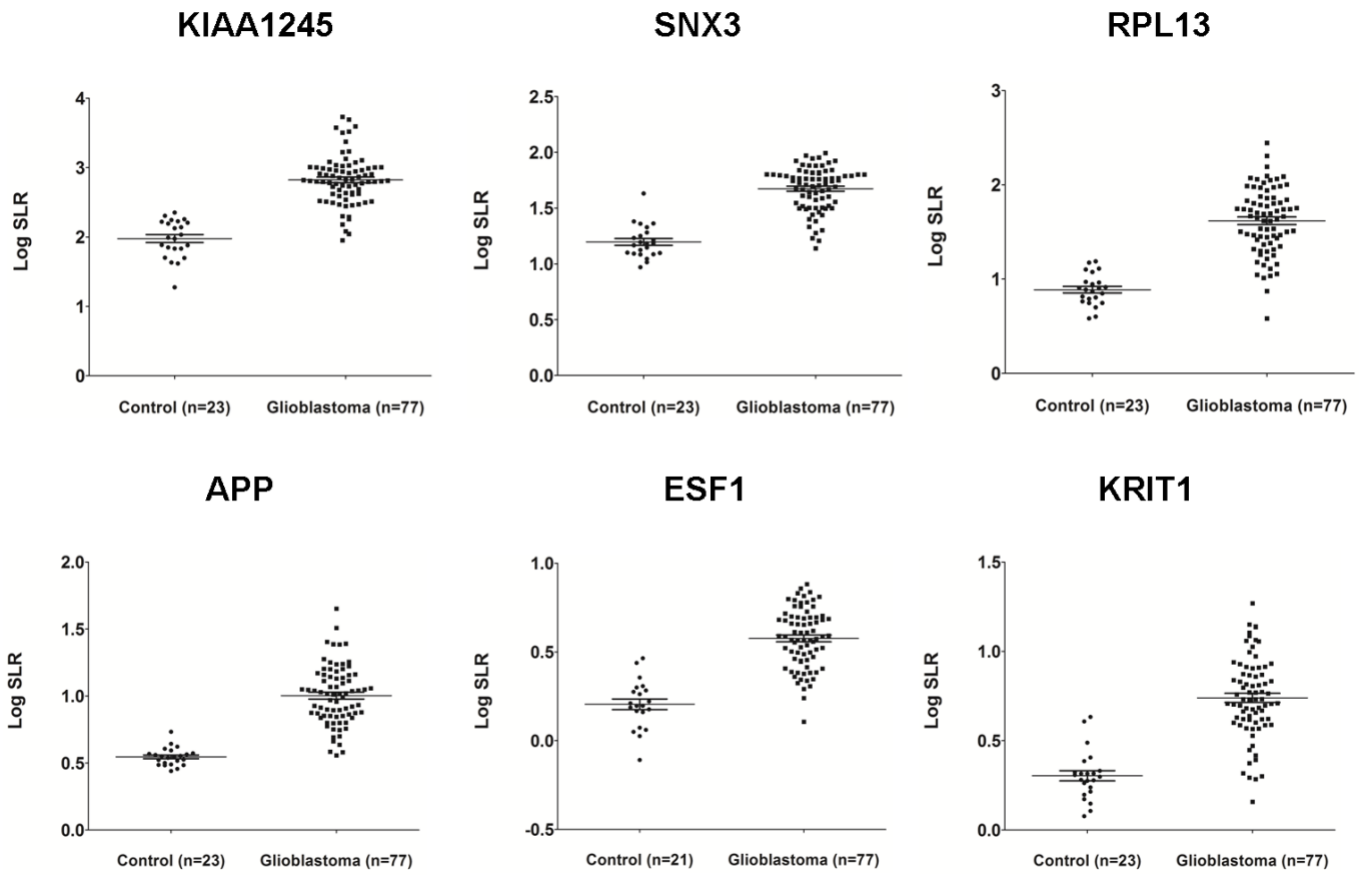


Figure 3.14 Scatter plot of the 6 genes that undergoes most significant 3'UTR shortening in Glioblastoma patients (n=77) as compared to normal samples (n=23) in GSE4290. The scatter plot illustrates the genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

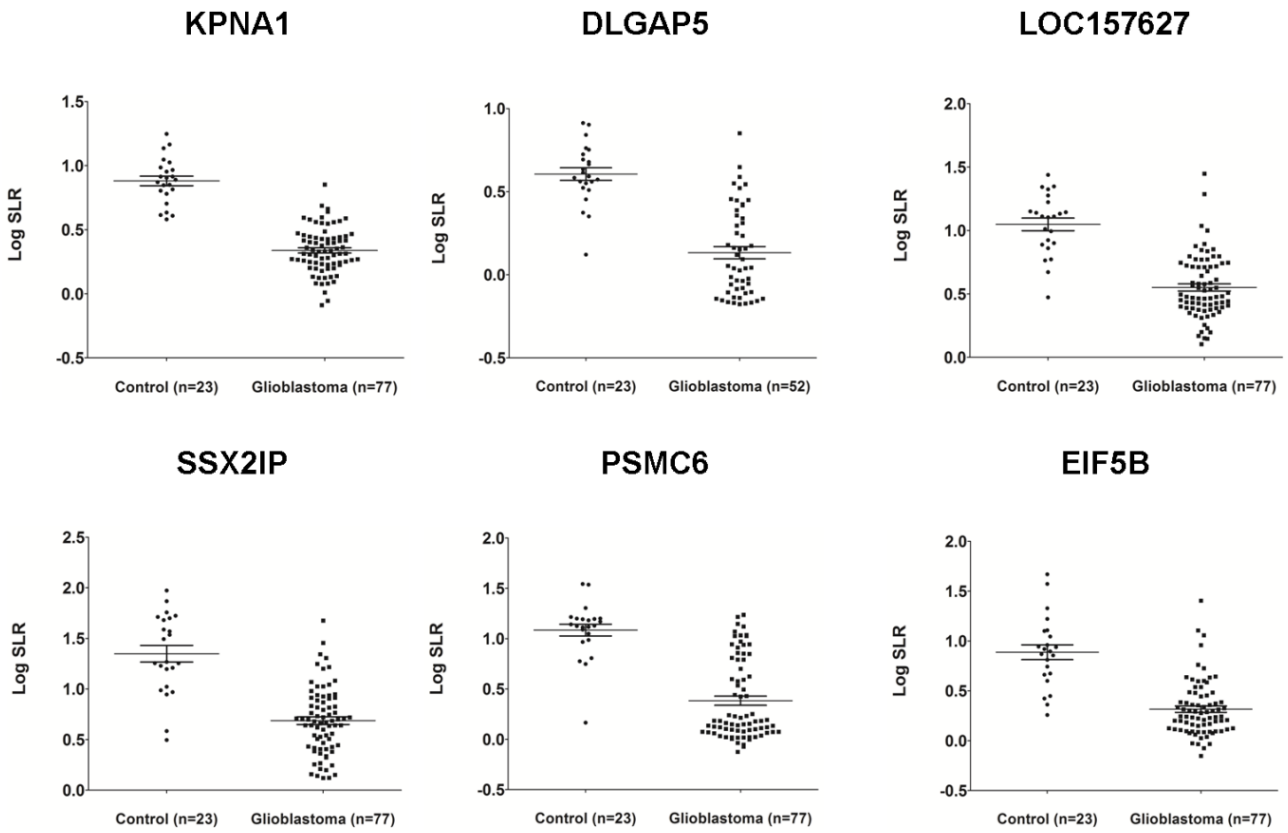


Figure 3.15 Scatter plots of the 6 genes that undergo significant 3'UTR lengthening in Glioblastoma tumors (n=77) as compared to normal tissue (n=23). The scatter plots illustrates the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

The scatter plots in Figures 3.14 and 3.15 clearly show that the genes with significantly different SLR values according to SAM, really do deviate largely from healthy controls in comparison. Table 3.6 gives detailed information about these genes' functions.

Table 3.6 Genes undergoing 3' UTR shortening/lengthening in Glioblastoma dataset GSE4290

Gene	Function/Relevance to Cancer	References
Shortening Events		
KIAA1245 (Neuroblastoma breakpoint family, member 12)	-duplicated gene family -linked to developmental and neurogenetic diseases	97
SNX3 (Sorting nexin 3)	-sorting nexin protein family -intracellular trafficking	73
RPL13 (Ribosomal protein L13)	-60S ribosomal subunit component	98
APP (Amyloid beta (A4) precursor protein)	-cell surface receptor -transmembrane precursor -linked to Alzheimer's disease	99
ESF1 (ESF1, nucleolar pre-rRNA processing protein, homolog (S. cerevisiae))	-potential transcription regulator by interaction with TATA-binding proteins	100
KRIT1 (KRIT1, ankyrin repeat containing)	-localized to cytoplasm and nucleus -role in maintenance of cell junctions, beta-1-integrin-mediated cell proliferation	101
Lengthening Events		
KPNA1 (Karyopherin alpha 1 (importin alpha 5))	-nuclear protein import mediator	102
DLGAP5 (Discs, large (Drosophila) homolog-associated protein 5)	-potential oncogene -upregulation in liver, prostate cancers	103 104
LOC157627 (LINC00599 long intergenic non-protein coding RNA 599)	-long non-coding RNA	105

Table 3.6 (continued)

SSX2IP (Synovial sarcoma, X breakpoint 2 interacting protein)	-centrosome formation -microtubule assembly -associated with leukemia, liver cancer	106 107
PSMC6 (Proteasome (prosome, macropain) 26S subunit, ATPase, 6)	-proteasome subunit -ATPase	108
EIF5B (Eukaryotic translation initiation factor 5B)	-accuracy in translation initiation -translational GTPase	109

3.2.2 APA in GSE16011 Dataset

This dataset contained a total of 284 samples, 8 of which were non-tumor samples from normal adult brain. 159 of the glioma samples were primary glioblastoma tumor samples. Therefore, the analysis and comparison were carried out on the 8 control and 159 tumor samples.

After APADetect analysis, firstly, tumor tissue samples were compared to control samples in SAM (Figure 3.16).

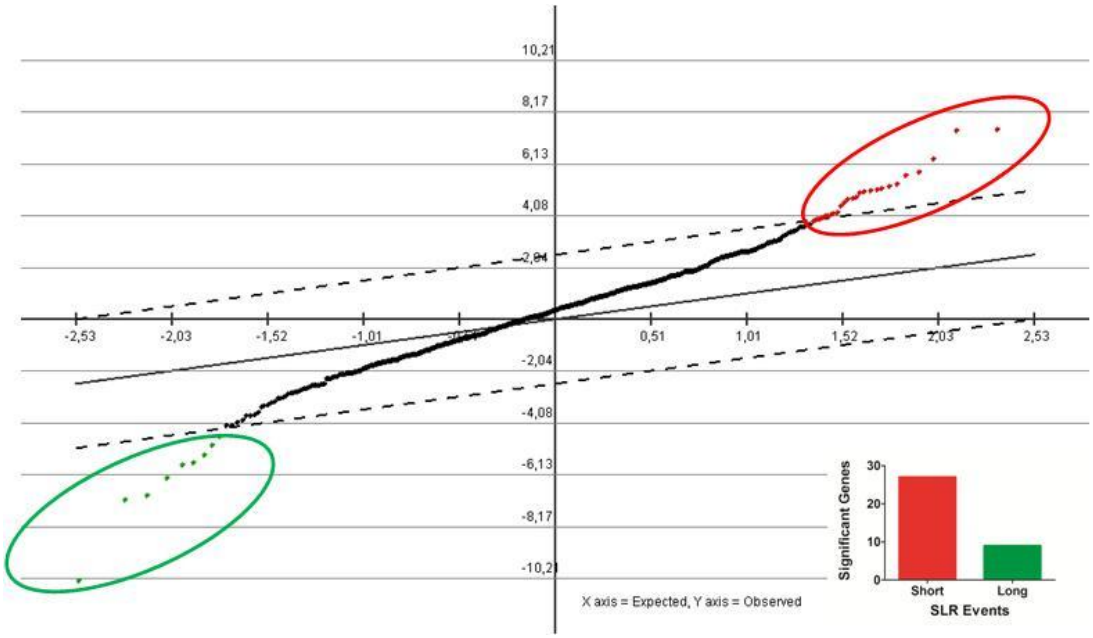


Figure 3.16 SAM graph of normal samples (n=8) vs. Glioblastoma tumor samples (n=159) in GSE16011. In the comparison between normal samples and tumor, the red and green dots indicate shortened and lengthened gene transcripts, respectively. In this case, there were 27 shortening and 9 lengthening events. All significant genes were identified as statistically significant with q-values<2%.

SAM results indicate 27 shortened and 9 lengthened transcripts (Appendix Tables B.9 and B.10). Scatter plots were prepared for the 6 most significantly shortened (*TCF3, METTL9, DNPEP, ZAK, SNX3 and TOP2A*) and 6 lengthened genes (*NEFH, SSX2IP, NMT2, LOC157627, KIF5A and SECISBP2L*) using individual data obtained from APADetect (Figures 3.17 and 3.18).

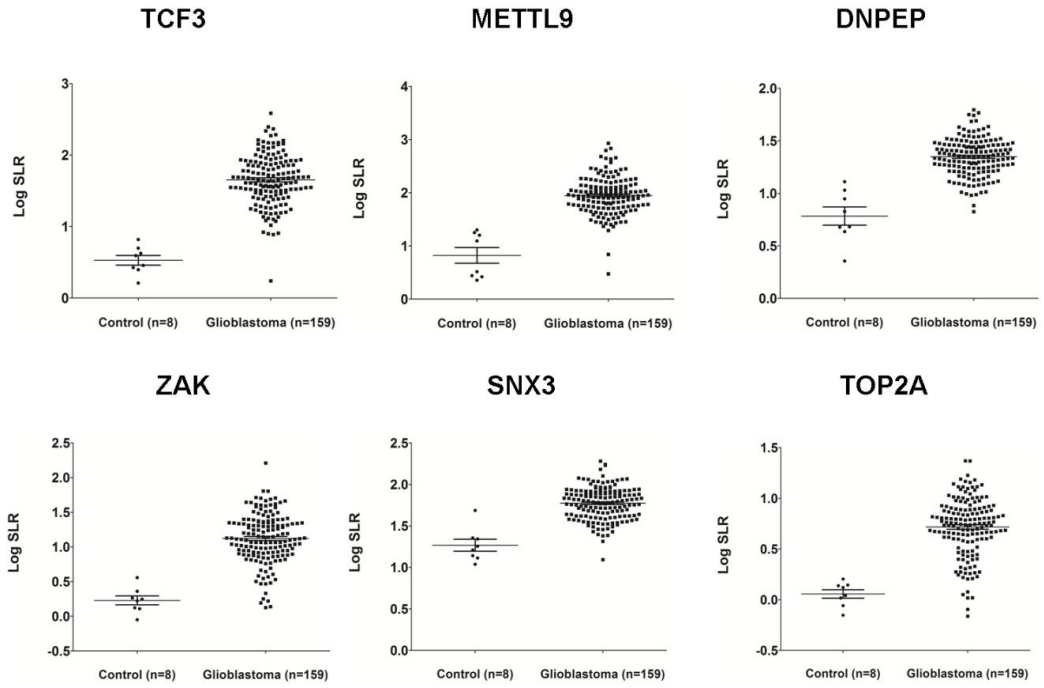


Figure 3.17 Scatter plot of the 6 genes that undergoes significant 3'UTR shortening in Glioblastoma patients (n=159) as compared to normal samples (n=8) in GSE16011. The scatter plot illustrates the genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

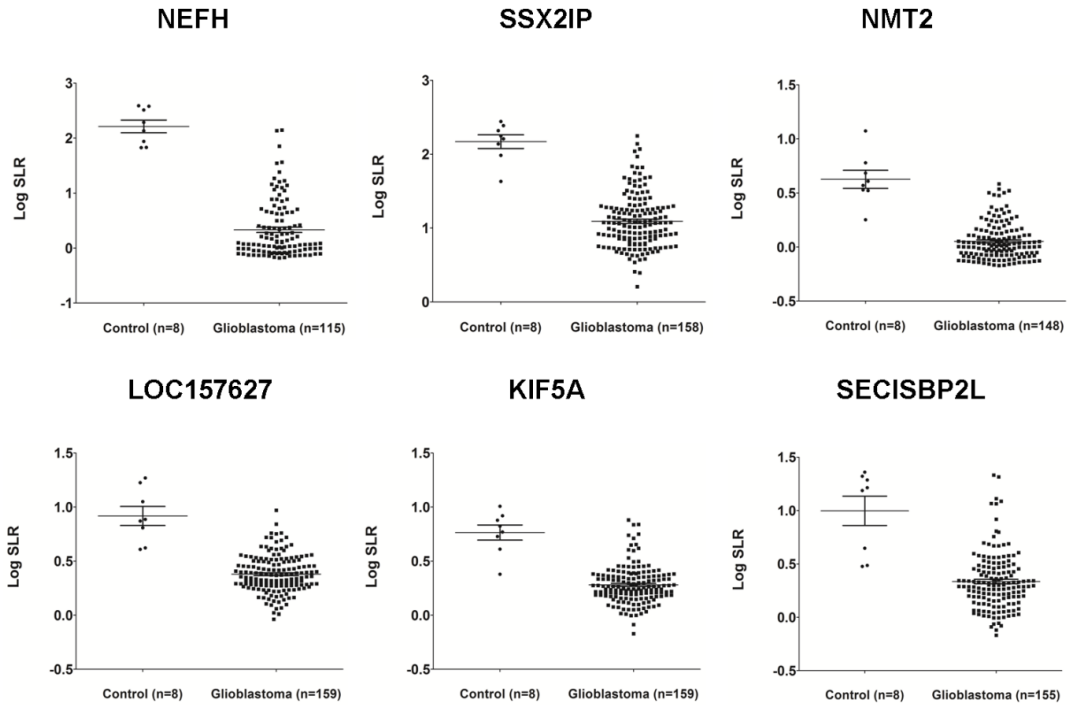


Figure 3.18 Scatter plots of the 6 genes that undergo significant 3'UTR lengthening in Glioblastoma tumors (n=159) as compared to normal tissue (n=8). The scatter plots illustrates the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

The scatter plots in Figures 3.17 and 3.18 clearly show that the genes with significantly different SLR values according to SAM, really do deviate largely from healthy controls in comparison. Table 3.7 gives detailed information about these genes' functions.

Table 3.7 Genes undergoing 3' UTR shortening/lengthening in Glioblastoma dataset GSE16011

Gene	Function/Relevance to Cancer	References
Shortening Events		
TCF3 (Transcription factor 3)	-helix-loop-helix transcription factor -crucial in lymphopoiesis -suppressor of Wnt pathway	110 111
METTL9 (Methyltransferase like 9)	-also known as *PAP1* -immunoglobulin superfamily protein -rat homologue: *IGSF6*	112
DNPEP (Aspartyl aminopeptidase)	-cytosolic acidic aminopeptidase -potential role in blood pressure maintenance or ocular functions	113 114
ZAK (Sterile alpha motif and leucine zipper containing kinase AZK)	-MAPKKK family protein -cell cycle checkpoint regulator -pro-apoptotic activity	86
SNX3 (Sorting nexin 3)	-sorting nexin protein family -intracellular trafficking	73
TOP2A (Topoisomerase (DNA) II alpha 170kDa)	-nuclease/ligase -ATP-dependent -DNA structure maintenance	115
Lengthening Events		
NEFH (Neurofilament, heavy peptide)	-type IV intermediate filament -neuronal cytoskeletal element -downregulation in renal, prostate cancer	116 117
SSX2IP (Synovial sarcoma, X breakpoint 2 interacting protein)	-centrosome formation -microtubule assembly -associated with leukemia, liver cancer	106 107 118
NMT2 (N-myristoyltransferase 2)	-co/post-translational modifier enzyme -crucial for growth -anti-cancer therapy target	119
LOC157627 (LINC00599 long intergenic non-protein coding RNA 599)	-long non-coding RNA	105
KIF5A (kinesin family member 5A)	-kinesin heavy chain -intracellular organelle transport	120
SECISBP2L (SECIS binding protein 2-like)	-selenocysteine insertion sequence binding protein -role in selenocysteine incorporation using UGA stop codon	121

Although there were significant differences between the two datasets that we analyzed, both datasets also showed similar outcomes (Figure 3.19).

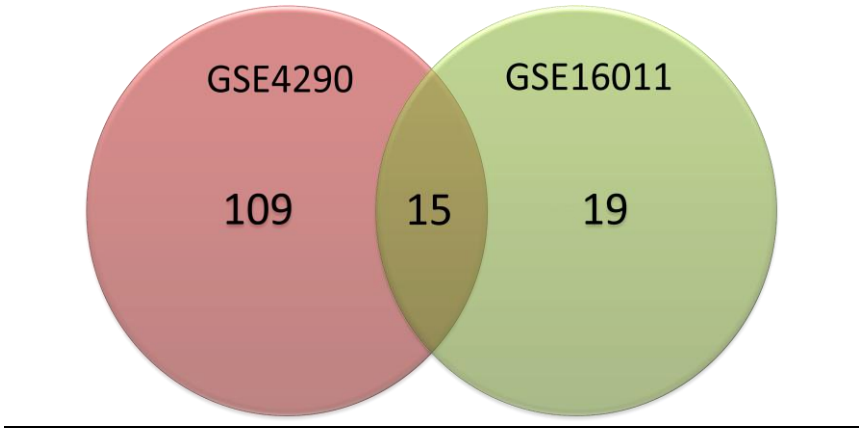


Figure 3.19 Venn diagram of significantly altered genes in the two datasets. Out of all significant alterations 15 genes are commonly altered in the two datasets.

The significance of these commonalities between the datasets was determined using Fisher's Exact Test. Our datasets showed significant overlap in common APA event changes with a p-value of 0.0012. This means that the APA changes in the 15 significantly altered genes are specific to glioblastoma tumor tissues and independent of differential expression between different brain regions and/or different microarray experiments.

In addition to APADetect and SAM analyses ontology search was also performed for genes with significantly altered APA patterns (Figure 3.20).

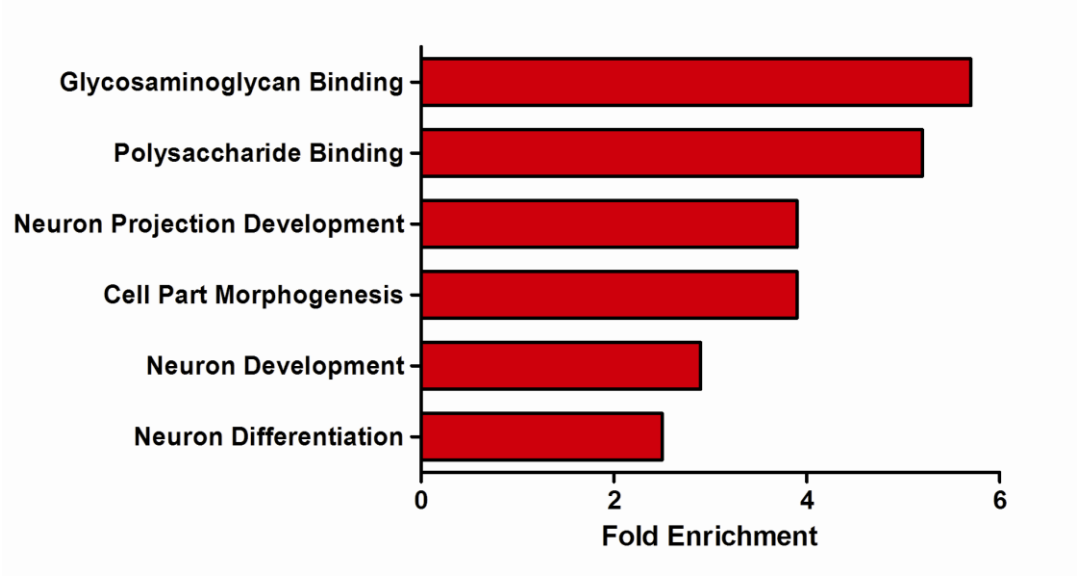


Figure 3.20 Enrichment graph of 142 genes showing significantly altered APA in glioblastoma datasets in terms of both shortening and lengthening. Gene ontology search was carried out for all the significantly altered genes in glioblastoma in terms of APA. DAVID results show enrichment in genes having functions related to neural differentiation and development. The gene clusters were obtained using Functional Annotation Clustering Tool. The most significant cluster had an enrichment score of 2.42, containing 7-8 polysaccharide-binding protein genes. The second most significant cluster had an enrichment score of 2.23, containing 9-10 neuron differentiation and development genes. Enrichment scores were obtained using the high stringency filter of the software. (p-value<0.001 for each cluster)

The enrichment scores of the clusters are statistically significant, and the p-values for each enriched cluster were less than 0.001. As can be seen from Figure 3.20, around 6-fold increase is seen in transcripts that show glycosaminoglycan binding. This may be especially important, considering that the brain structure and the network of brain cell connections are highly complex. Upon tumor formation, new links and attachments are likely to be needed and this may be leading to the significant enrichment that we observe in polysaccharide-binding transcripts. Also, significantly altered transcripts are enriched in that an increase of at least 2-fold is seen in genes

having important functions in neuron differentiation and development. This indicates that important tissue specific APA changes can occur in proliferative disease states in addition to genes that have roles in cell division cycle. APA in cell cycle related genes' expression are most likely to be affected by abnormal rapid divisions, but differential APA changes can also be observed in a tissue specific manner. This is another aspect on the vastness of changes that take place in the cells as they undergo malignant transformations.

In order to check the validity of our approach, scatter plots were prepared for genes whose SLR values were equal to one. As expected, no significant difference between control samples and patient samples could be observed (Figure 3.21).

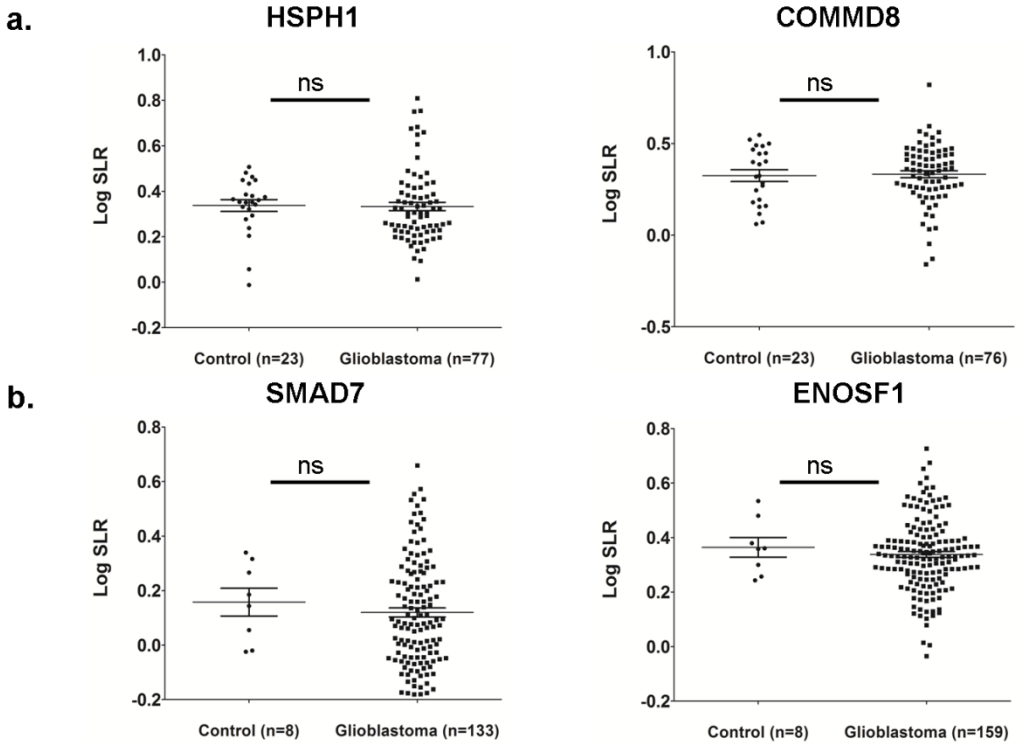


Figure 3.21 Scatter plots of genes that do not undergo 3'UTR alteration in (a.) GSE4290 and (b.) GSE16011 as compared to normal samples. The scatter plots illustrate 4 genes that showed no significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

3.2.3 APA in GSE4271 Dataset

In this dataset 23 primary high grade glioma tumor samples were analyzed and compared to their matched recurrences in paired analysis to find out whether recurrence would be associated with an altered APA pattern using APADetect and SAM (Figure 3.22).

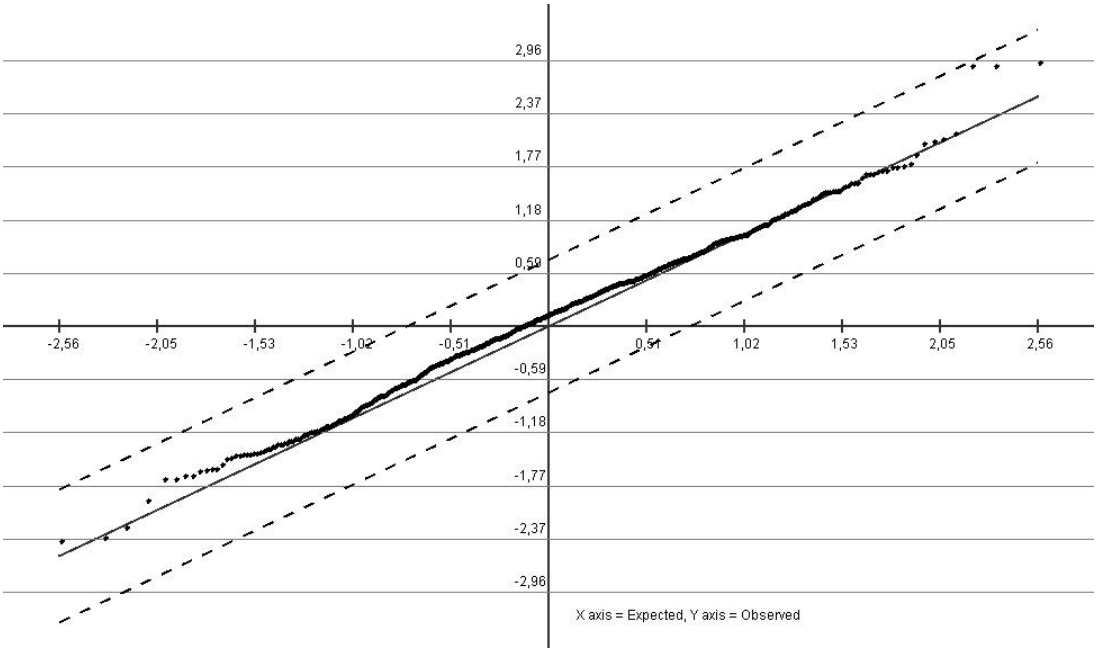


Figure 3.22 SAM graph of tumor samples (n=23) vs. matched recurrence samples (n=23). In the paired comparison between tumor samples and their matched recurrences, there are no significantly altered gene transcripts.

SAM results indicate no shortening or lengthening event changes when tumor samples and their matched recurrences were compared. This could mean that APA is not significantly altered as the disease shows recurrence. Therefore, we think APA pattern changes may occur as early events during neoplastic transformations.

3.3 APA in Gastric Cancer

Gastric cancer is another disease that cannot be diagnosed early and easily because its symptoms are not very distinctive, leading to metastasis and low overall survival [46]. Therefore, APA pattern changes may be used to help diagnosis of this disease. For APA analysis in gastric cancer, we used GSE29272 dataset [58]. This dataset contained 62 gastric cardia adenocarcinoma tissues with their matched normal samples and 72 gastric noncardia adenocarcinoma tissues with their matched normal samples. Gastric cardia tissue is the region which is in close proximity of the opening that connects the stomach to the esophagus, whereas gastric noncardia tissue includes the rest of the stomach.

3.3.1 APA in Gastric Cardia and Noncardia Adenocarcinomas

To detect gastric cancer specific APA patterns, cardia (n=62) and noncardia (n=72) samples were analyzed with their respective matched controls (n=62 and n=72) using APADetect and SAM (Figure 3.23).

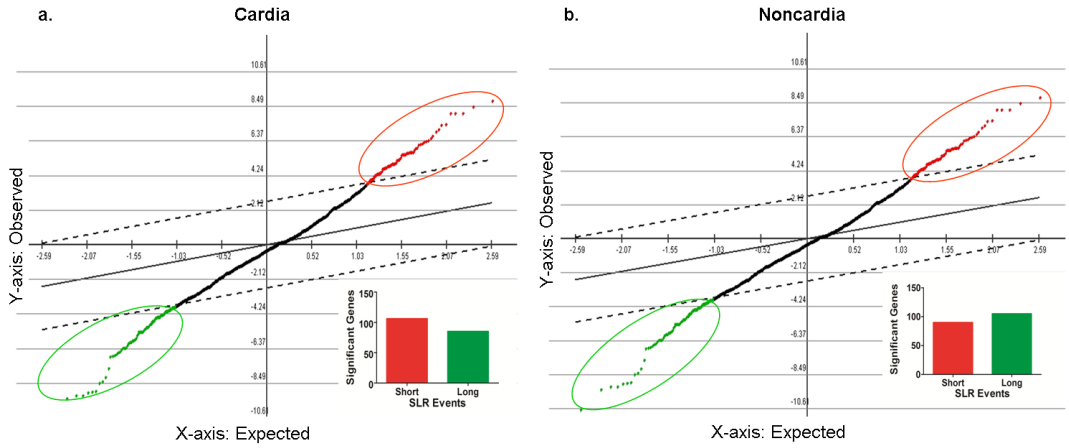


Figure 3.23 SAM graph of (a.) cardia tumor samples (n=62) compared to matched controls (n=62) and (b.) noncardia tumor samples (n=72) compared to matched controls (n=72). In the comparisons the red and green dots indicate shortened and lengthened gene transcripts, respectively. In cardia tumors, 106 genes undergo significant shortening while 85 genes undergo significant lenghtening. In noncardia tumors, 90 genes undergo significant shortening while 105 genes undergo significant lengthening. All significant genes were identified as statistically significant with q-values<2%.

SAM results indicate 106 shortened and 85 lengthened transcripts in cardia while there appears to be 90 shortened and 105 lengthened transcripts in noncardia group of patients (Appendix Tables B.11-B.14). Scatter plots were prepared for the 6 most significantly shortened and 6 lengthened genes using individual data obtained from APADetect for both cardia (Figures 3.24 and 3.25), and noncardia group of patients (Figures 3.26 and 3.27).

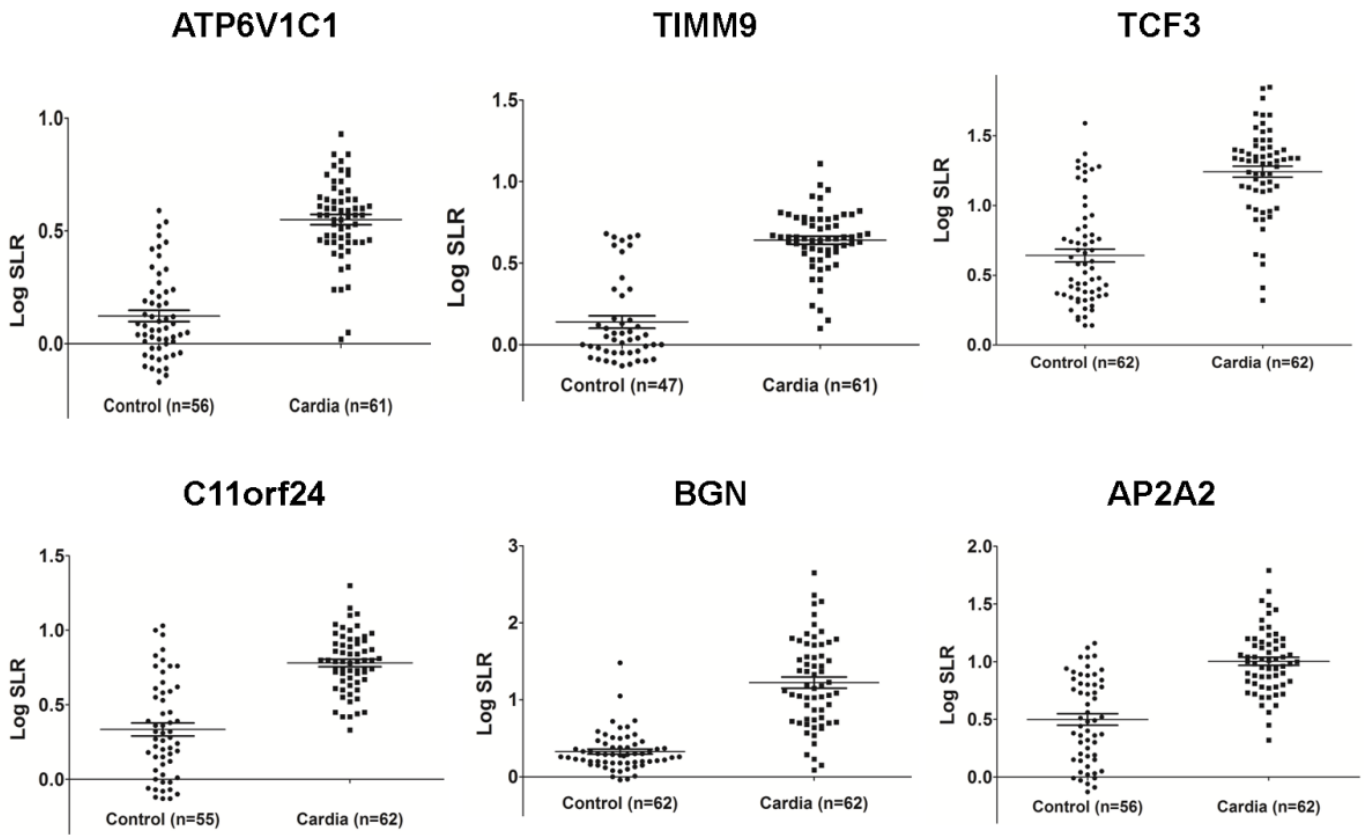


Figure 3.224 Scatter plots of 6 genes that undergo significant 3'UTR shortening in cardia adenocarcinoma (n=62) compared to normal tissue (n=62). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

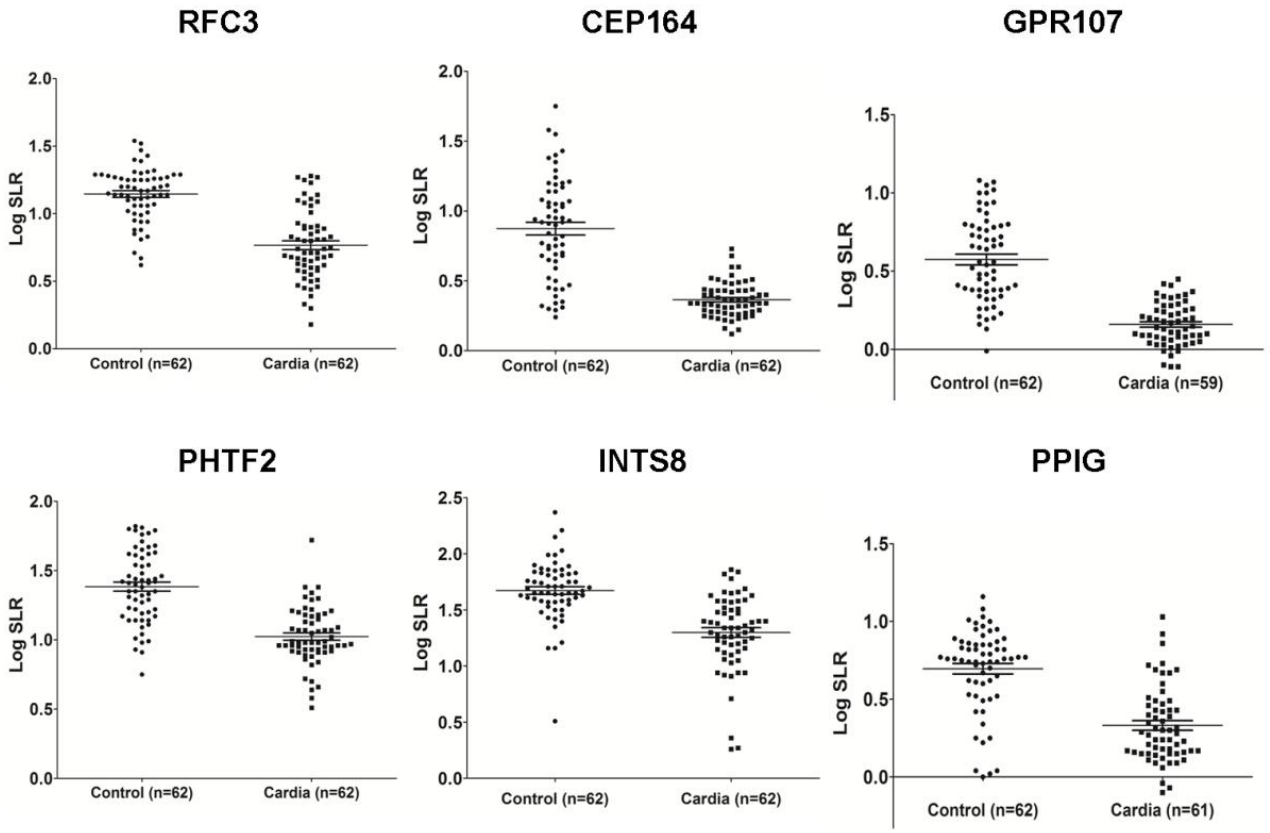


Figure 3.35 Scatter plots of the 6 genes that undergo significant 3'UTR lengthening in cardia adenocarcinoma (n=62) compared to normal tissue (n=62). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

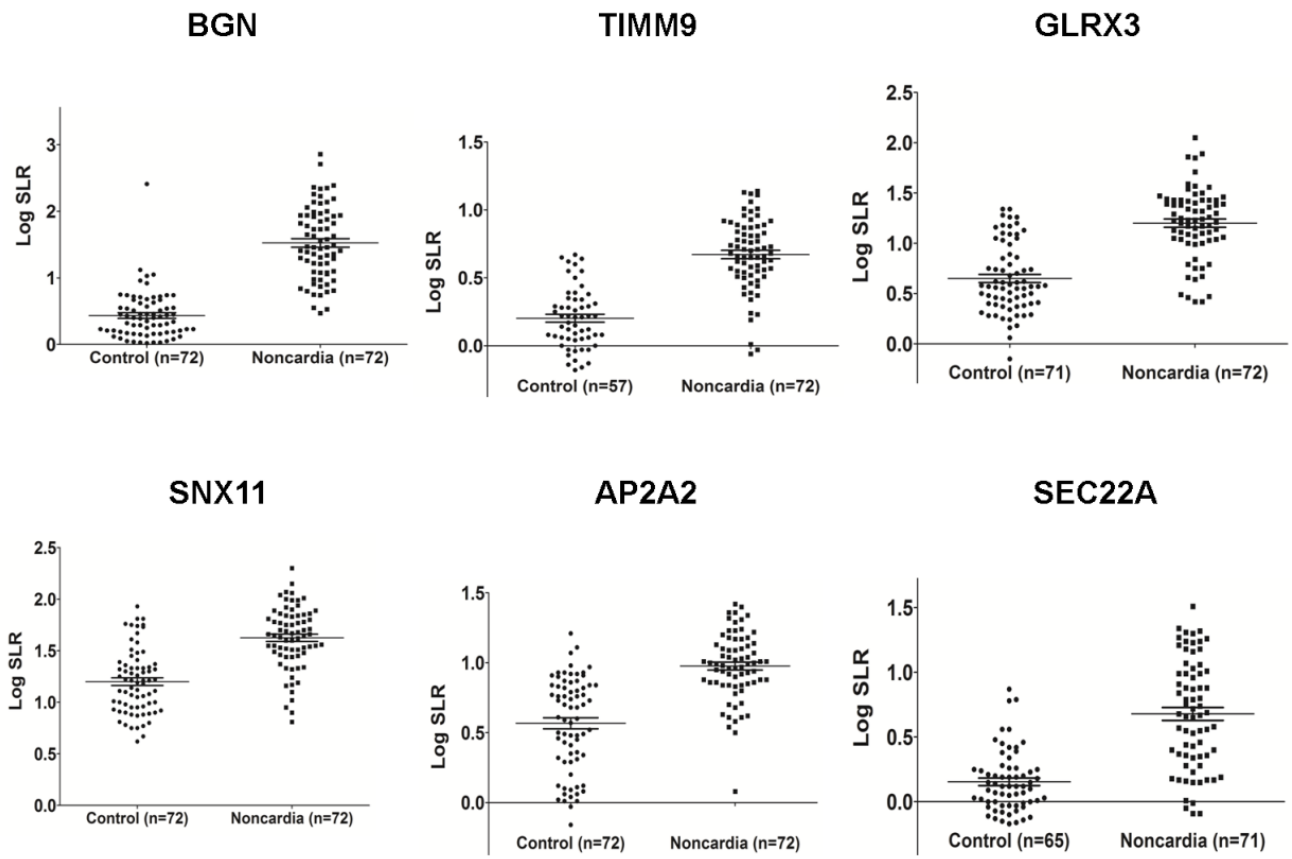


Figure 3.26 Scatter plots of 6 genes that undergo significant 3'UTR shortening in noncardia adenocarcinoma (n=72) compared to normal tissue (n=72). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

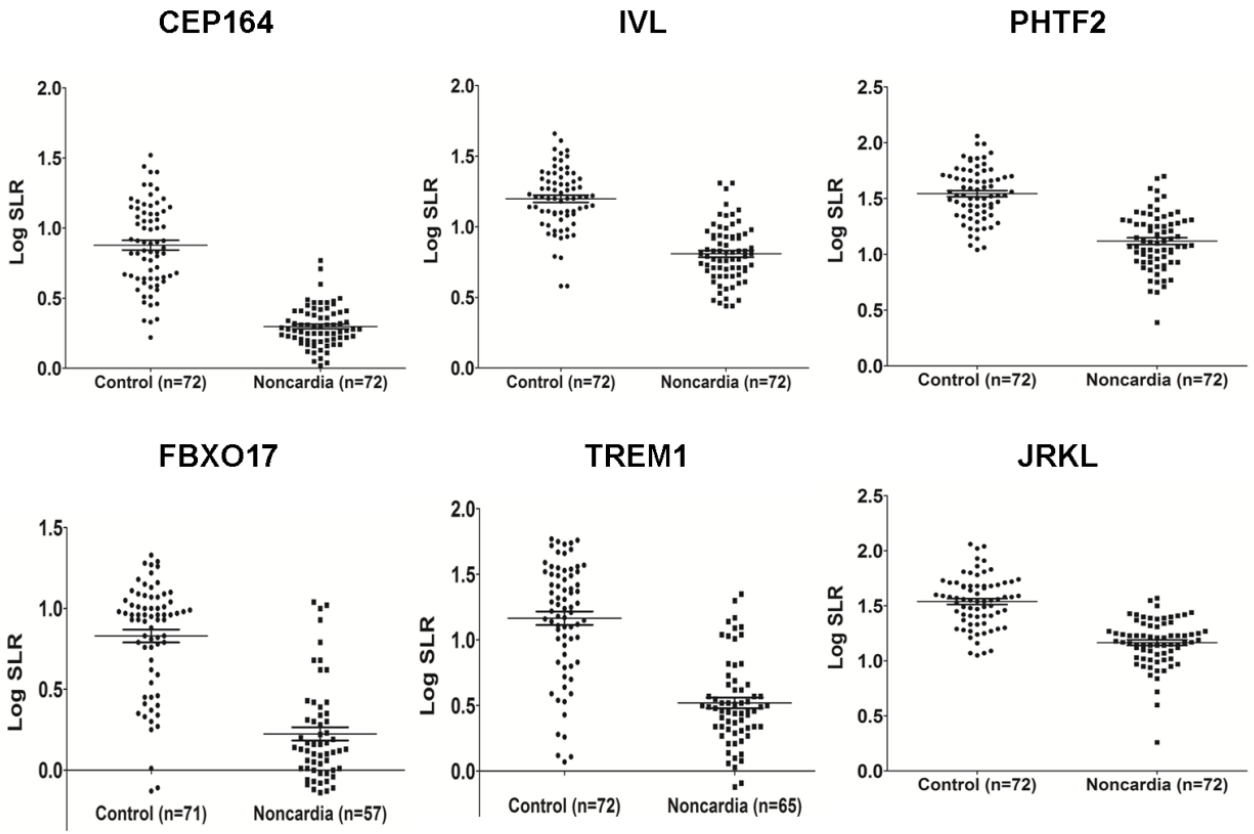


Figure 3.27 Scatter plots of the 6 genes that undergo significant 3'UTR lengthening in noncardia adenocarcinoma (n=72) compared to normal tissue (n=72). The scatter plots illustrate the 6 genes that showed highest significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

The scatter plots through Figures 3.23 - 3.26 clearly show that the genes with significantly different SLR values according to SAM, really do deviate largely from healthy controls in comparison. Table 3.8 gives detailed information about these genes' functions.

Table 3.8 A selection of genes undergoing 3' UTR shortening/lengthening in cardia and noncardia tumors

Gene	Function/Relevance to Cancer	References
Shortening Events		
ATP6V1C1 (ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C1)	-vacuolar ATPase component -thought to be important in cytoskeletal arrangement -correlated with high metastatic capability of breast cancer	122 123
TIMM9 (Translocase of inner mitochondrial membrane 9 homolog)	-mitochondrial biogenesis/translation marker -elevated expression in epithelial breast cancer cells	124
TCF3 (Transcription factor 3)	-helix-loop-helix transcription factor -crucial in lymphopoiesis -suppressor of Wnt pathway	110 111
C11orf24 (Chromosome 11 open reading frame 24)	-thought to be a Golgi protein	125
BGN (Biglycan)	-small, leucine-rich proteoglycan -important extracellular matrix component -role in bone, tendon development -may be related to metastatic capability	126 127 128 129
AP2A2 (Adaptor-related protein complex 2, alpha 2 subunit)	-subunit of α-Adaptin -related to clathrin-mediated endocytosis -involved in immune system related functions	130 131
GLRX3 (Glutaredoxin 3)	-oxidoreductase enzyme -potential apoptosis inhibitor -potential cancer marker	77
SNX11 (Sorting nexin 11)	-sorting nexin protein family -intracellular trafficking	73
SEC22A (SEC22 vesicle trafficking protein homolog A)	-member of SEC22 protein family functioning in vesicular trafficking -potential role in secretory pathway	132 133

Table 3.8 (continued)

Lengthening Events		
RFC3 (Replication factor C (activator 1) 3, 38 kDa)	-critical subunit of RFC with role in elongation of primed DNA during replication -role in DNA damage response -potential oncogene	134 135
CEP164 (Centrosomal protein 164 kDa)	-centriole marker -ciliogenesis -potentially key component in ciliary mambrane biogenesis	136 137
GPR107 (G protein-coupled receptor 107)	-orphan receptor -candidate neuronostatin (may be initiating specific signaling in gastric cancer) receptor	138
PHTF2 (Putative homeodomain transcription factor 2)	-undefined	
INTS8 (Integrator complex subunit 8)	-role in processing of small nuclear RNAs U1 and U2 -association with RNA polymerase II -upregulated in gastric tumor tissues	139
PPIG (Peptidylprolyl isomerase G (cyclophilin G))	-nuclear cyclophilin -spliceosome component	140
IVL (Involucrin)	-keratinocyte crosslinked envelope component -transglutaminase substrate -potentially important in epidermal integrity	141 142
FBXO17 (F-box protein 17)	-F-box protein family member -glycoprotein metabolism -ubiquitination	143 144
TREM1 (Triggering receptor expressed on myeloid cells 1)	-expression in neutrophils and monocytes -inflammatory response mediator	145
JRKL (JRK-like)	-undefined	

Another point of interest was whether there were differences in APA patterns between cardia and noncardia tumors due to anatomical site. After APADetect analysis, cardia and noncardia tumor tissue samples were normalized to their respective control samples. Then, SAM was performed and significantly different genes were scatter-plotted (Figures 3.28 and 3.29).

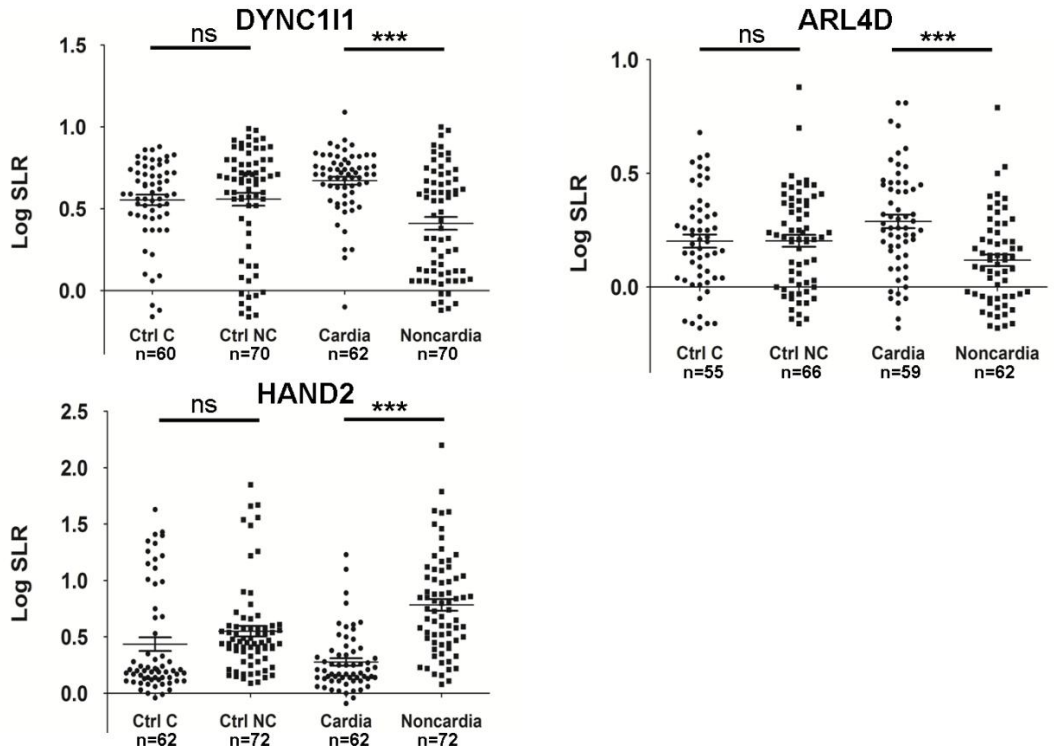


Figure 3.28 Scatter plots of three genes that undergo significantly different 3'UTR processing in cardia (n=62) and noncardia tumors (n=72). The scatter plots illustrate the genes that showed significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization. "ns" indicates "not significant" in that there is no statistical difference between normal cardia and noncardia samples. "***" indicates highly significant difference between cardia and noncardia tumor samples with a p-value of less than 0.001, according to both SAM and ANOVA.

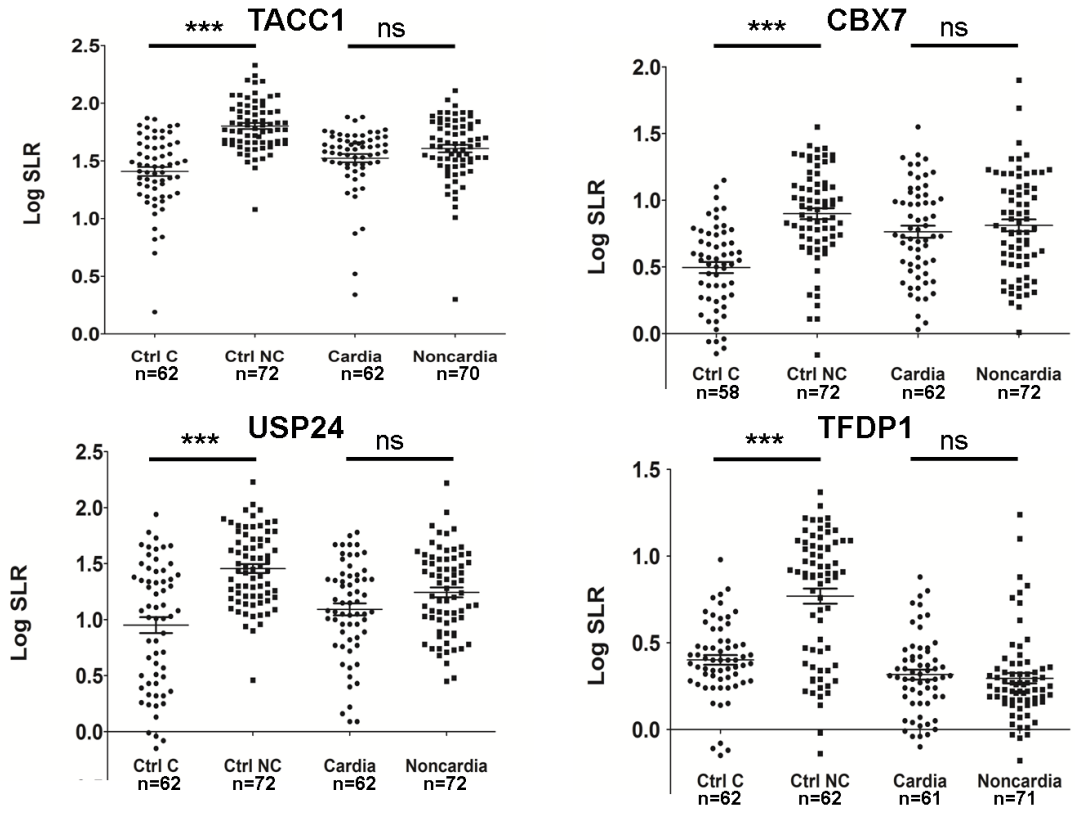


Figure 3.29 Scatter plots of the genes that undergo significantly different 3'UTR processing in between normal cardia (n=62) and noncardia regions of the stomach (n=72). The scatter plots illustrate the genes that showed significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization. "ns" indicates "not significant" in that there is no statistical difference between cardia and noncardia tumor samples. "***" indicates highly significant difference between normal cardia and noncardia samples with a p-value of less than 0.001, according to both SAM and ANOVA.

The scatter plots in Figures 3.27 and 3.28 illustrate that the genes with significantly different SLR values according to SAM, really do deviate largely between cardia and noncardia normal tissue or cancerous tissue suggesting certain APA pattern differences resulting from anatomical site differences. According to Figure 3.27,

while there was no APA-based change in controls of both groups, there were significant changes in cardia and noncardia tumors for the *TACC1, CBX7, USP24* and *TFDP1* genes. On the contrary, in Figure 3.28, while normal controls had different APA patterns, cardia and noncardia samples were not significantly different. These possible anatomical site specific alterations will be interesting to validate in terms of functional and biological relevance. Table 3.9 gives detailed information about these genes' functions.

Table 3.9 Genes undergoing 3' UTR shortening/lengthening in an anatomical site difference manner in cardia and noncardia samples

Gene	Function/Relevance to Cancer	References
Shortening Events in Cancer Tissue		
DYNC1I1 (Dynein, cytoplasmic 1, intermediate chain 1)	-neuronal retrograde axonal transporter	146
ARL4D (ADP-ribosylation factor-like 4D)	-ADP-ribosylation factor -potential role in intracellular trafficking -nuclear localization signal presence	147
Lengthening Event in Cancer Tissue		
HAND2 (Heart and neural crest derivatives expressed 2)	-helix-loop-helix transcription factor -differentially methylated in endometrial cancer	148
Shortening Events in Normal Tissue		
TACC1 (Transforming, acidic coiled-coil containing protein 1)	-centrosome/microtubule binding protein -potential links to gastric cancer, breast cancer, glioblastoma	149 150 151
CBX7 (Chromobox homolog 7)	-repressor complex component -upregulated in lymphoma -downregulated in various malignancies -potentially oncogenic	152 153
USP24 (Ubiquitin specific peptidase 24)	-cystein protease -deubiquitinating enzyme	154

Table 3.9 (continued)

TFDP1 (Transcription factor Dp-1)	-heterodimerization with E2F proteins -transcriptional regulator of cell cycle proteins	155

Out of all the shortening and lengthening events in cardia and noncardia tumors, 56 genes were commonly shortened while 50 genes were commonly lengthened (Figure 3.30 - Appendix Table B.15).

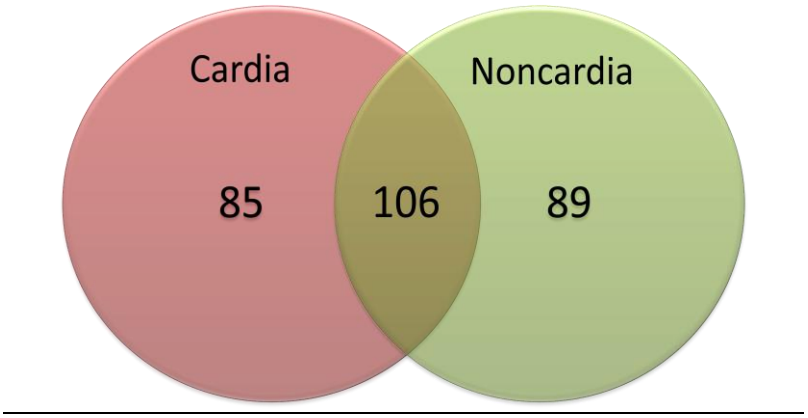


Figure 3.30 Venn diagram of significantly altered genes in cardia and noncardia tumors. Out of all significant alterations 106 genes are commonly altered in cardia and noncardia tumors.

Fisher's Exact Test was performed in order to check whether these commonly altered genes are statistically significant. Fisher's Exact Test results indicate high statistical

significance for these overlaps with a p-value of less than 0.0001. Observation of such common changes in cancer states of the two different anatomical sites belonging to the same tissue could be used to obtain an idea about the progression of cardia and noncardia tumors. These altered genes could be contributing to the malignant transformation directly or they could be affected from the transformation indirectly. Either way they could be useful in identification of the disease.

In addition to APADetect and SAM analyses, ontology search was also performed for genes with significantly altered APA patterns in cardia (Figure 3.31) and noncardia (Figure 3.32) tumor samples.

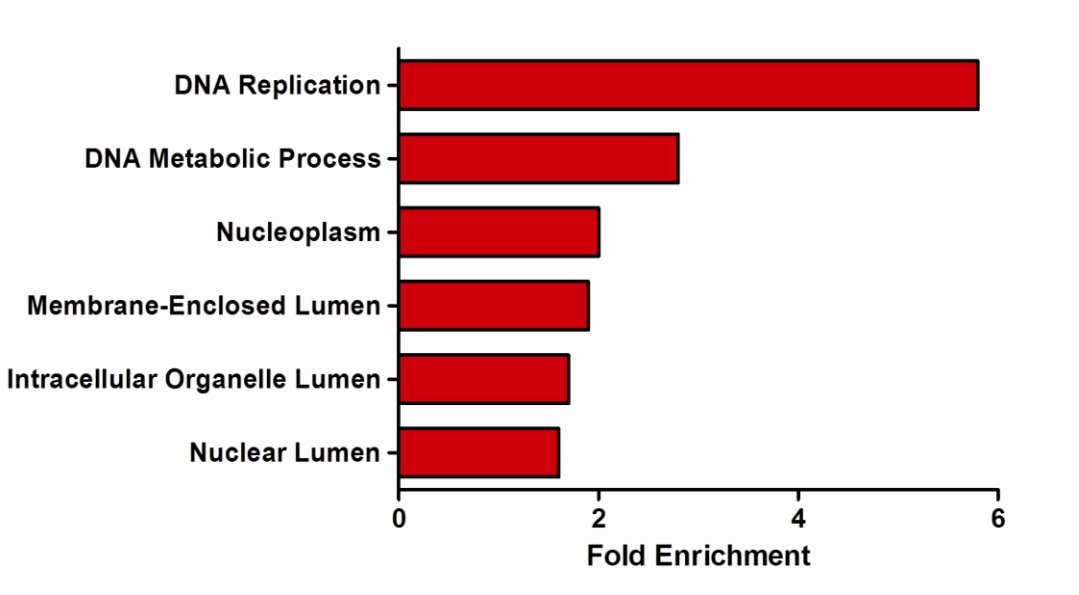


Figure 3.31 Enrichment graph of 164 genes showing significantly altered APA in cardia in terms of both shortening and lengthening. Gene ontology search was carried out for all the significantly altered genes in cardia tumors in terms of APA. DAVID results show significant enrichment in genes having functions related to DNA replication and metabolism, and in turn cell division. The gene clusters were obtained using Functional Annotation Clustering Tool. The most significant cluster had an enrichment score of 3.83, containing 7-14 DNA replication/metabolism genes. The second most significant cluster had an enrichment score of 2.52, containing 17-34 organelle or nuclear lumen protein genes. Enrichment scores were obtained using the medium stringency filter of the software. (p-value<0.001 for each cluster)

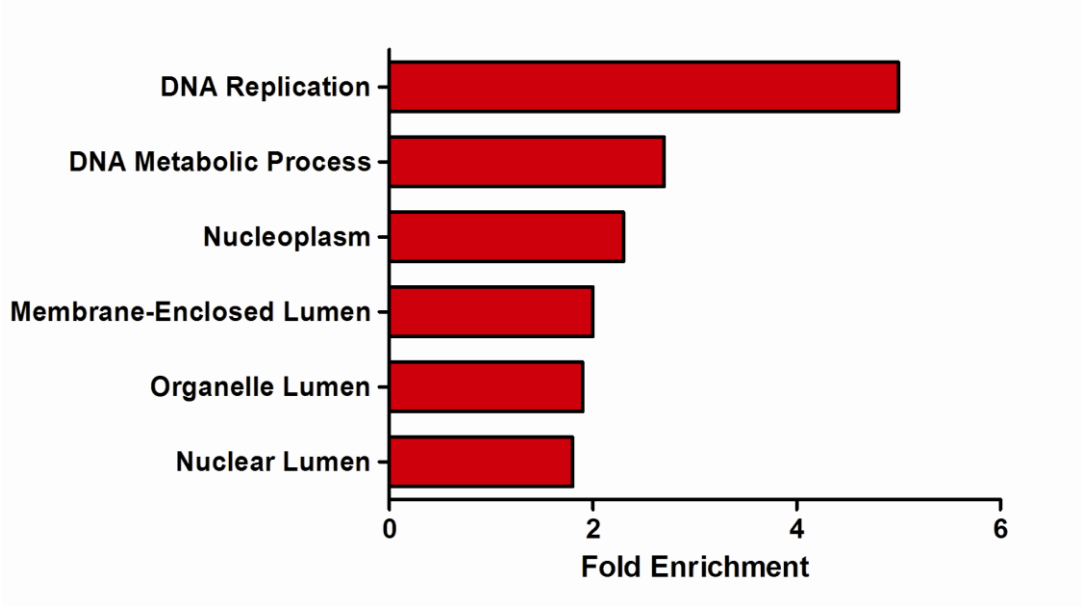


Figure 3.32 Enrichment graph of 173 genes showing significantly altered APA in noncardia in terms of both shortening and lengthening. Gene ontology search was carried out for all the significantly altered genes in noncardia tumors in terms of APA. DAVID results show significant enrichment in genes having functions related to DNA replication and metabolism, and in turn cell division. The gene clusters were obtained using Functional Annotation Clustering Tool. The most significant cluster had an enrichment score of 3.13, containing 6-14 DNA replication/metabolism genes. The second most significant cluster had an enrichment score of 3.02, containing 13-38 organelle or nuclear lumen protein genes. Enrichment scores were obtained using the medium stringency filter of the software. (p-value<0.001 for each cluster)

The enrichment scores of the clusters are statistically significant, and the p-values for each enriched cluster were less than 0.001. As can be seen from Figures 3.31 and 3.32, significantly altered transcripts are enriched in that an increase of around 6-fold is seen in genes that have functions in DNA replication. Also an increase of up to 2-fold can be seen in genes located in membrane-bound lumens, especially nucleoplasm. This is consistent with the enrichment of transcripts that are important

in DNA metabolism. This is another indication of how APA pattern changes can be related to progression of proliferative and cancerous diseases.

For control of the validity of this approach, scatter plots were prepared for genes whose SLR values were equal to one. As expected, no significant difference between control samples and patient samples could be observed (Figure 3.33).

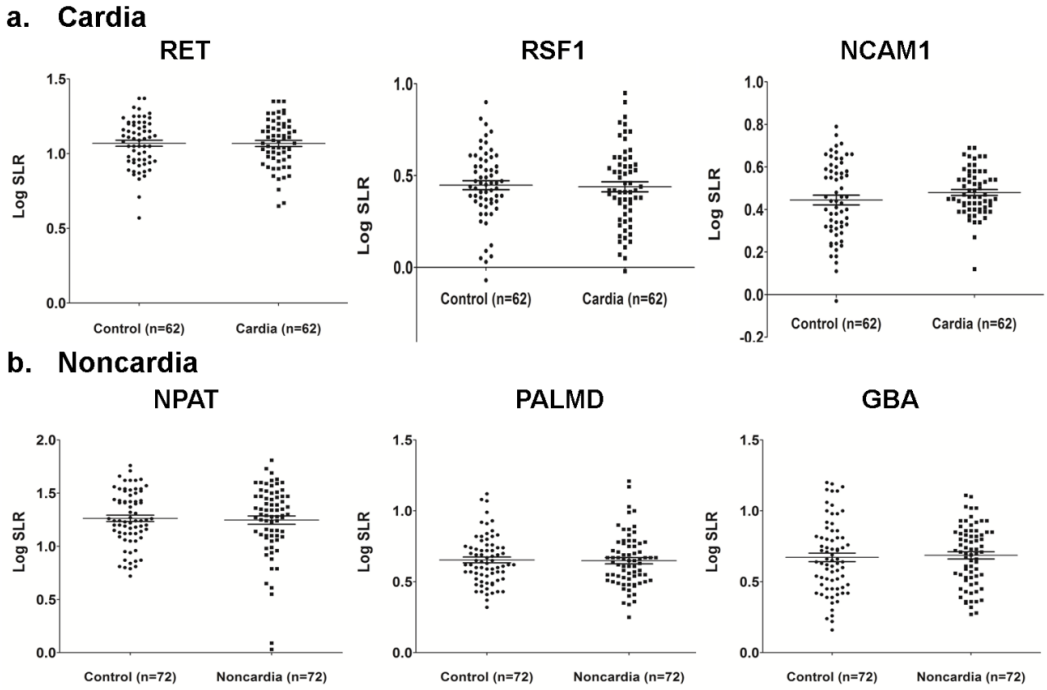


Figure 3.33 Scatter plots of genes that do not undergo 3'UTR alteration in (a.) cardia and (b.) noncardia as compared to normal samples. The scatter plots illustrate 6 genes that showed no significance after APADetect and SAM analyses. Log transformed SLR values were used in plots for clear visualization.

3.3.2 APA Patterns in Relation to Patient Characteristics

Another approach used to extend the knowledge about cardia and noncardia tumors was to check whether APA pattern changes significantly correlate with patient characteristics such as survival, gender and family history of gastrointestinal cancer. Because gastric cancer cannot easily be diagnosed in early stages [46], most of the samples were high grade, so whether tumor grade is related with APA changes could not be tested.

There were no significantly different alternatively polyadenylated genes when males were compared to females, or when patients with no family of GI cancer were compared to patients with family history of GI cancer. Future studies with higher numbers of patients may be more suitable for such correlative studies.

Survival analysis was also carried out between high SLR patients and low SLR patients. The high and low SLR patients correspond to the top and bottom ~35% of the patients in terms of SLR values. High SLR indicates 3' UTR shortening while low SLR is closer to or even less than normal SLR. Out of 25 genes, only *TOP2A* had a correlation with survival of patients in cardia tumors. *TOP2A* survival curve is given in Figure 3.34.

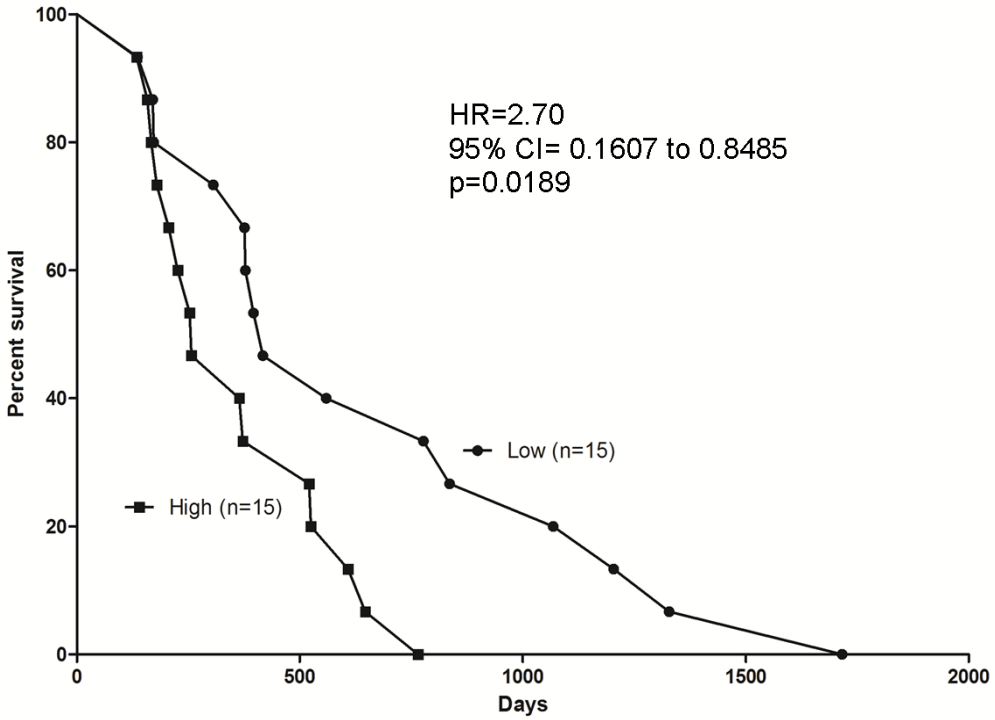


Figure 3.34 Survival curve of *TOP2A* gene. Survival curve of *TOP2A* high and low SLR patients (n=15 each) for a period of 5 years (1865 days). Hazard ratio for high SLR patients for *TOP2A* was 2.7 (95% CI, 0.1607-0.8485).

Interestingly, while the survival rates are low in gastric cancer, *TOP2A* SLRs (High vs Low) had a correlation (Hazard Ratio of 2.7 (95% CI, 0.1607-0.8485)) with the survival times in the cardia patients (Figure 3.32). *TOP2A* and cancer connection is already being investigated. Both its gene copy number and expression levels are known to be changed in cancer cells [156]. *TOP2A* is located at chromosome 17q [157], and alterations in this region are commonly seen especially in the form of amplifications in many different cancers such as gastric, breast, ovarian or bladder

cancers [157-159]. Therefore, screening higher numbers of patients for *TOP2A* SLR and survival connection could potentially be important.

Overall, the data presented here shows that diverse APA pattern changes can occur in proliferative/cancerous diseases. The changes differ not only for solid tumors and hematologic malignancies, but also for solid tumors of different tissue types, while intriguing examples of commonalities exist such as *TCF3*, *GLRX3, SNX3* or *SEC22A*.

3.4 Experimental Confirmation of an APA Regulated Candidate Gene

Given the information we have generated on different cancers, and how APA may be a regulator of gene expression, we were interested to verify shortening of *TCF3*. *TCF3* appeared to be 2.78 fold shortened in glioblastoma, 1.78 fold shortened in gastric cardia cancer, and 1.36 fold shortened in gastric noncardia cancer.

3' RACE was carried out, followed by cloning and sequencing for *TCF3* gene in order to confirm the presence of short transcripts, experimentally. *TCF3* was among the most significantly shortened genes, it appeared to be alternatively polyadenylated in different tumor types and it could potentially be important in progression of these proliferative diseases due to its function. *TCF3* gene encodes the transcription factor 3 protein, also named *E2A*. It belongs to the helix-loop-helix transcription factor family, and is known to be especially important in myeloid malignancies [110,111]. Because *TCF3* is associated with suppresion of Wnt signalling pathway [160], and over activation of Wnt signalling is associated with cancer [161], differential APA patterns of *TCF3* may be important in the pathology of proliferative diseases. Therefore, 3' RACE specific primers were designed to verify that short 3' UTR isoform of *TCF3* indeed existed (Figure 3.35).

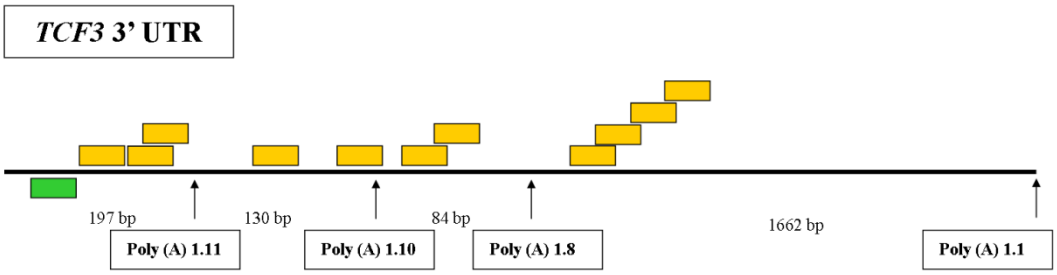
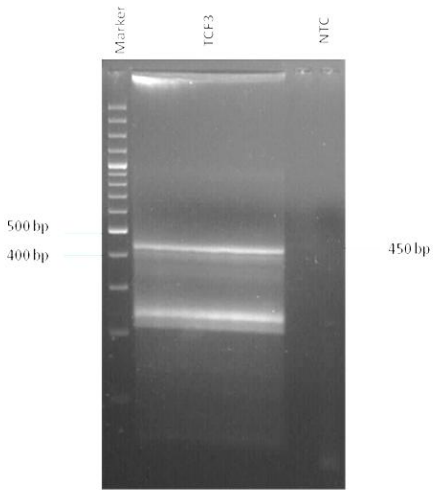


Figure 3.35 Structure of *TCF3* transcript's 3' UTR. *TCF3* contains 4 different polyadenylation sites on the transcript according to polyA_DB [66]. The yellow boxes indicate Affymetrix probe locations and the green box indicates *TCF3* forward primer location. The reverse anchor primer is from the anchor sequence that is attached to oligo(dT)$_{18}$ primer for 3' RACE specific cDNA preparation. The indicated sizes are from the beginning of the forward primer to the marked poly(A) sites.

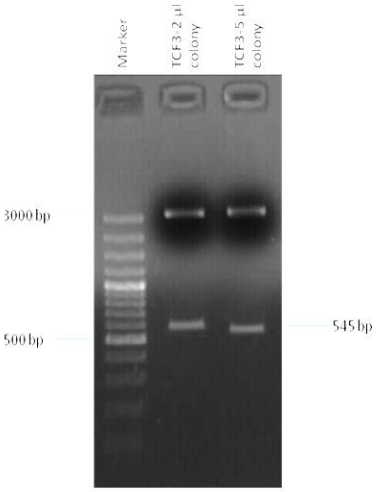
According to APADetect analysis, the transcript isoform that is polyadenylated from Poly(A) 1.8 position is the one that is upregulated. Based on the positions of the *TCF3* forward primer and reverse anchor primer for 3' RACE, the expected size of the transcriptionally upregulated short transcript isoform was 450 bp (from forward primer to poly(A) 1.8 and oligo(dT)$_{18}$ with anchor sequence). Indeed the primers were optimized to generate the expected size of 450 bp (Figure 3.36a).

Next, the PCR products were purified from agarose gel to be cloned into pGEM-T Easy Vector System (Promega). Following transformation into XL-1 Blue cells, antibiotic selected cells were grown to check whether they carried the plasmids with the desired inserts. For both inserts vectors were double digested with Apa I and Sac I at 37^{o}C for 2 hours. Results confirmed the correct insert size for 2 constructs (Figure 3.36b). Sequencing results also confirmed the poly(A) tail positions, showing the presence of the short isoform (Figure 3.36c-d).

a.



b.



c.

```
TCF3   2400 GNCCAANCCTTTGTCAGAAAGAAAAATGCCTTACTATAAAAAGC
TCF3   2444 GGAGAAATGGAAACATATCACTCAAGGGGGATGCTGTGGAAACC
TCF3   2488 TGGCTTATTCTTCTAAAGCCACCAGCAAATTGTGCCTAAGCGAA
TCF3   2532 ATATTTTTTTTTAAGGAAAATAAAAACATTAGTTACAAGATTTTT
TCF3   2576 TTTTTCTTAATGTAGATGAAAATTAGCAAGGATGCTGCCTTTGG
TCF3   2620 TCTCTGGTTTTTTTAAGCTTTTTTTGCATATGTTTTGTAAGCAA
TCF3   2664 CAAATTTTTTTGTATAAAAGTCCCGTGTCTCTCGCTATTTCTGC
TCF3   2708 TGCTGTTCCTAGACTGAGCATTGCATTTCTTGATCAACCAGATG
TCF3   2752 ATTAAACGTTGTATTAAAAAAAAAAAAAAAAAAAAAAAAAAA
                                  Poly(A) Tail
TCF3   2796 AAAAAAAAAAAAAAAAAAAAAAGTCGACATCGATACGCGTGGT
```

d.

```
TCF3   2399 GGGCNNCCTTTGTCAGAAAGAAAAATGCCTTACTATAAAAAGCG
TCF3   2443 GAGAAATGGAAACATATCACTCAAGGGGGATGCTGTGGAAACCT
TCF3   2487 GGCTTATTCTTCTAAAGCCACCAGCAAATTGTGCCTAAGCGAAA
TCF3   2531 TATTTTTTTTTAAGGAAAATAAAAACATTAGTTACAAGATTTTTT
TCF3   2575 TTTTCTTAATGTAGATGAAAATTAGCAAGGATGCTGCCTTTGGT
TCF3   2619 CTCTGGTTTTTTTAAGCTTTTTTTGCATATGTTTTGTAAGCAAC
TCF3   2663 AAATTTTTTTGTATAAAAGTCCCGTGTCTCTCGCTATTTCTGCT
TCF3   2707 GCTGTTCCTAGACTGAGCATTGCATTTCTTGATCAACCAGATGA
TCF3   2751 TTAAACGTTGTATTAAAAAAAAAAAAAAAGTCGACATCGATACG
                                  Poly(A) Tail
TCF3         CGTGGT
```

Figure 3.36 3' RACE, double digestion and sequencing of *TCF3*. a. 3’ RACE result for the shorter 3’ UTR isoform of *TCF3*. Expected result for the gene specific forward and 3’RACE specific anchor reverse primer was 450 bp. b. ApaI and SacI digest of pGEM-T_TCF3 construct confirmed the cloning of the RACE product. Insert size was 450 bp, digested fragment of 545 bp contained additional vector sequence. Two colonies were chosen for sequencing due to the different cloning product size. c-d. Sequencing of the inserts of the pGEM-T confirmed the existence of the proximal poly(A) tail and the short 3’ UTR isoform. Sequencing result showed the size difference to be due to poly(A) tail length.

Sequencing of the two different pGEM-T_TCF3 plasmid constructs proved the presence of short *TCF3* transcript of the expected size in AGS gastric cancer cell line.

CHAPTER 4

CONCLUSION

Alternative polyadenylation is the process by which transcript isoforms of different lengths are generated in the cell [162]. Because it is an important regulation mechanism, significant alterations in APA patterns are likely to be associated with different disease states. It is already known that rapidly proliferating cells tend to generate shorter transcripts [163]. This may result in loss of microRNA (miRNA) binding sites on the transcript and may confer an escape route from negative regulatory mechanisms involving 3' UTRs. Such an arrangement could increase translational efficiency of the mRNA and oncogenic expression might be favored [164].

Myelodysplastic syndromes (MDSs), glioblastoma multiforme (GBM), and gastric cancers [153] are aggressive proliferative diseases. Early diagnoses of these malignancies often exhibit serious challenges, and cannot be performed easily. As a result, these diseases manifest poor prognosis and poor survival rates. For development of solutions to these problems, it is of utmost importance to define the molecular events underlying the pathology of such diseases. Considering that alternative polyadenylation regulation is important, its alterations are likely to have major impact on disease progression. Therefore, we wanted to investigate alterations in APA patterns in MDS, GBM and GC patients as compared to their respective healthy control groups.

For this purpose, we analyzed microarray data of patient sets retrieved from NCBI GEO using APADetect and SAM. As a result of SAM, a high number of genes have been found to be significantly altered in disease states in terms of alternative polyadenylation. Most significantly altered genes were plotted on scatter plots for better visualization of the alteration and it was interesting to note that these most significantly altered genes were either already associated with cancer, or they had important functions that are potentially important in disease progression. Another interesting point was that while 3' UTR shortening can be observed in potentially oncogenic transcripts, 3' UTR lengthening could be seen in potentially tumor suppressive gene transcripts.

We also wanted to investigate whether enrichment patterns could be observed in significantly altered genes' functions. Therefore, we performed gene ontology search using DAVID. As a result we found that genes with significant alterations in APA patterns show enrichment in functions either related to cell division cycle or tissue specific processes. This finding is supportive of the knowledge that APA is associated with cancer.

Another point of interest was to find out if APA pattern changes could be associated with patient survival. We obtained detailed patient information containing survival status of gastric cancer patients [58] and performed survival analysis for high SLR (Short/Long Ratio) and low SLR patients. Survival analysis did not reveal much significance in terms of APA and survival relation. However, only for *TOP2A* gene, high SLR seemed to be correlating with lower survival. Such an observation is important in that, this gene could potentially be used as a prognostic indicator in the future. It may also indicate that APA pattern changes are occurring as early events during tumorigenesis. Therefore, this result seems promising, but verification of this correlation in larger patient groups is necessary still.

Finally, we wanted to verify our findings experimentally. To this end, we performed 3' RACE (Rapid Amplification of cDNA Ends) for *TCF3* gene transcript. The reason why *TCF3* was of interest was that it seemed to undergo significant APA change in

different cancer types including breast cancer (data not shown) and could potentially be important in disease progression. Cloning and sequencing results confirmed our bioinformatics analyses' results and verified the presence of transcript isoform with shorter 3' UTR, in a gastric cancer cell line.

In future prospect, these findings may be important for identification of novel diagnostic/prognostic markers. Examination of commonalities between these diverse diseases may point out to a common mechanism underlying the pathology of proliferative diseases. In addition to bioinformatics, further experimental verifications and confirmations of these results in larger datasets is necessary before such a suggestion can be made. Hence, APA analyses in other patient datasets may be performed to find out if the results overlap with our findings. In addition 3' RACE may be used to confirm transcript isoform presence for other potentially important genes. Also, relative expressions of these isoforms may be measured using RT-qPCR via isoform specific primers. To conclude, in light of these findings and along with future studies, we hope to contribute to the knowledge of the interplay between alternative polyadenylation and proliferative disease pathology.

REFERENCES

1. Murugan, R. & Kreiman, G. Theory on the coupled stochastic dynamics of transcription and splice-site recognition. *PLoS Comput Biol* **8**, e1002747 (2012).

2. Lutz, C.S. & Moreira, A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip Rev RNA* **2**, 23-31 (2011).

3. Kuhn, U. *et al.* Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J Biol Chem* **284**, 22803-14 (2009).

4. Eckmann, C.R., Rammelt, C. & Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* **2**, 348-61 (2011).

5. Elkon, R., Ugalde, A.P. & Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**, 496-506 (2013).

6. Griseri, P. & Pages, G. Regulation of the mRNA half-life in breast cancer. *World J Clin Oncol* **5**, 323-34 (2014).

7. Proudfoot, N.J. Ending the message: poly(A) signals then and now. *Genes Dev* **25**, 1770-82 (2011).

8. Hollerer, I., Grund, K., Hentze, M.W. & Kulozik, A.E. mRNA 3'end processing: A tale of the tail reaches the clinic. *EMBO Mol Med* **6**, 16-26 (2014).

9. Di Giammartino, D.C., Nishida, K. & Manley, J.L. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**, 853-66 (2011).

10. Tian, B., Hu, J., Zhang, H. & Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* **33**, 201-12 (2005).

11. Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M. & Gautheret, D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* **10**, 1001-10 (2000).

12. Sun, Y., Fu, Y., Li, Y. & Xu, A. Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. *J Mol Cell Biol* **4**, 352-61 (2012).

13. Erlitzki, R., Long, J.C. & Theil, E.C. Multiple, conserved iron-responsive elements in the 3'-untranslated region of transferrin receptor mRNA enhance binding of iron regulatory protein 2. *J Biol Chem* **277**, 42579-87 (2002).

14. Takagaki, Y. & Manley, J.L. Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol Cell* **2**, 761-71 (1998).

15. Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A. & Handa, H. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Res* **34**, 6264-71 (2006).

16. Licatalosi, D.D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-9 (2008).

17. Jenal, M. *et al.* The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**, 538-53 (2012).

18. Castelo-Branco, P. *et al.* Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* **24**, 4174-83 (2004).

19. Bava, F.A. *et al.* CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* **495**, 121-5 (2013).

20. Spies, N., Nielsen, C.B., Padgett, R.A. & Burge, C.B. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* **36**, 245-54 (2009).

21. Lian, Z. *et al.* A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. *Genome Res* **18**, 1224-37 (2008).

22. Buratowski, S. Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* **17**, 257-61 (2005).

23. Nagaike, T. *et al.* Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**, 409-18 (2011).

24. Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L. & Milcarek, C. Transcription elongation factor ELL2 directs immunoglobulin secretion in

plasma cells by stimulating altered RNA processing. *Nat Immunol* **10**, 1102-9 (2009).

25. Shi, Y. Alternative polyadenylation: new insights from global analyses. *Rna* **18**, 2105-17 (2012).

26. Pinto, P.A. *et al.* RNA polymerase II kinetics in polo polyadenylation signal selection. *Embo j* **30**, 2431-44 (2011).

27. Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**, 2757-74 (2010).

28. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664-8 (2010).

29. Gautheret, D., Poirot, O., Lopez, F., Audic, S. & Claverie, J.M. Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering. *Genome Res* **8**, 524-30 (1998).

30. Bednar, M. DNA microarray technology and application. *Med Sci Monit* **6**, 796-800 (2000).

31. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009).

32. Pickrell, J.K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768-72 (2010).

33. Fu, Y. *et al.* Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* **21**, 741-7 (2011).

34. Alt, F.W. *et al.* Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**, 293-301 (1980).

35. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. & Burge, C.B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643-7 (2008).

36. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* **106**, 7028-33 (2009).

37. Novak, J., Bienertova-Vasku, J., Kara, T. & Novak, M. MicroRNAs involved in the lipid metabolism and their possible implications for atherosclerosis development and treatment. *Mediators Inflamm* **2014**, 275867 (2014).

38. Tothova, Z., Steensma, D.P. & Ebert, B.L. New strategies in myelodysplastic syndromes: application of molecular diagnostics to clinical practice. *Clin Cancer Res* **19**, 1637-43 (2013).

39. Samiev, D., Bhatt, V.R., Armitage, J.D., Maness, L.J. & Akhtari, M. A primary care approach to myelodysplastic syndromes. *Korean J Fam Med* **35**, 111-8 (2014).

40. Larsson, C.A., Cote, G. & Quintas-Cardama, A. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol Cancer Res* **11**, 815-27 (2013).

41. Vardiman, J.W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-51 (2009).

42. Jovcevska, I., Kocevar, N. & Komel, R. Glioma and glioblastoma - how much do we (not) know? in *Mol Clin Oncol*, Vol. 1 935-941 (2013).

43. Wilson, T.A., Karajannis, M.A. & Harter, D.H. Glioblastoma multiforme: State of the art and future therapeutics. *Surg Neurol Int* **5**, 64 (2014).

44. Zhang, H., Lee, J.Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol* **6**, R100 (2005).

45. Jemal, A. *et al.* Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61**, 69-90.

46. Tsai, M.M. *et al.* Potential prognostic, diagnostic and therapeutic markers for human gastric cancer. *World J Gastroenterol* **20**, 13791-803 (2014).

47. Dhalla, F., da Silva, S.P., Lucas, M., Travis, S. & Chapel, H. Review of gastric cancer risk factors in patients with common variable immunodeficiency disorders, resulting in a proposal for a surveillance programme. *Clinical & Experimental Immunology* **165**, 1-7 (2011).

48. Kim, B. *et al.* Expression profiling and subtype-specific expression of stomach cancer. *Cancer Res* **63**, 8248-55 (2003).

49. Boussioutas, A. *et al.* Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. *Cancer Res* **63**, 2569-77 (2003).

50. Stomach cancer: MedlinePlus Medical Encyclopedia. (2014). (last accessed on 15.01.2015)

51. Kim, M.A., Lee, H.S., Yang, H.K. & Kim, W.H. Clinicopathologic and protein expression differences between cardia carcinoma and noncardia carcinoma of the stomach. *Cancer* **103**, 1439-46 (2005).

52. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**, D991-5 (2013).

53. Pellagatti, A. *et al.* Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* **24**, 756-64 (2010).

54. Sun, L. *et al.* Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* **9**, 287-300 (2006).

55. Gravendeel, L.A. *et al.* Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res* **69**, 9065-72 (2009).

56. Phillips, H.S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157-73 (2006).

57. Costa, B.M. *et al.* Reversing HOXA9 oncogene activation by PI3K inhibition: epigenetic mechanism and prognostic significance in human glioblastoma. *Cancer Res* **70**, 453-62 (2010).

58. Wang, G. *et al.* Comparison of global gene expression of gastric cardia and noncardia cancers from a high-risk population in china. *PLoS One* **8**, e63826 (2013).

59. Payton, J.E. *et al.* High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. *J Clin Invest* **119**, 1714-26 (2009).

60. McGowan, K.A. *et al.* Reduced ribosomal protein gene dosage and p53 activation in low-risk myelodysplastic syndrome. *Blood* **118**, 3622-33 (2011).

61. Akman, B.H., Can, T. & Erson-Bensan, A.E. Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Res* **40**, 10679-88 (2012).

62. Zhang, H., Hu, J., Recce, M. & Tian, B. PolyA_DB: a database for mammalian mRNA polyadenylation. *Nucleic Acids Res* **33**, D116-20 (2005).

63. Saeed, A.I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-8 (2003).

64. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-21 (2001).

65. Huang da, W. *et al.* DAVID gene ID conversion tool. *Bioinformation* **2**, 428-30 (2008).

66. Lee, J.Y., Yeh, I., Park, J.Y. & Tian, B. PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res* **35**, D165-8 (2007).

67. Malcovati, L. & Nimer, S.D. Myelodysplastic syndromes: diagnosis and staging. *Cancer Control* **15 Suppl**, 4-13 (2008).

68. Significance analysis of Microarrays: User guide and technical document - sam.pdf. (2014). (last accessed on 15.01.2014)

69. Yamayoshi, S., Ohka, S., Fujii, K. & Koike, S. Functional comparison of SCARB2 and PSGL1 as receptors for enterovirus 71. *J Virol* **87**, 3335-47 (2013).

70. Nepravishta, R., Bellomaria, A., Polizio, F., Paci, M. & Melino, S. Reticulon RTN1-C(CT) peptide: a potential nuclease and inhibitor of histone deacetylase enzymes. *Biochemistry* **49**, 252-8 (2010).

71. Wilkie-Grantham, R.P., Matsuzawa, S. & Reed, J.C. Novel phosphorylation and ubiquitination sites regulate reactive oxygen species-dependent degradation of anti-apoptotic c-FLIP protein. *J Biol Chem* **288**, 12777-90 (2013).

72. Cho, D.H. *et al.* Induced inhibition of ischemic/hypoxic injury by APIP, a novel Apaf-1-interacting protein. *J Biol Chem* **279**, 39942-50 (2004).

73. Xu, J. *et al.* Structure of sorting nexin 11 (SNX11) reveals a novel extended phox homology (PX) domain critical for inhibition of SNX10-induced vacuolation. *J Biol Chem* **288**, 16598-605 (2013).

74. Jeong, K.W. Flightless I (Drosophila) homolog facilitates chromatin accessibility of the estrogen receptor alpha target genes in MCF-7 breast cancer cells. *Biochem Biophys Res Commun* **446**, 608-13 (2014).

75. Yuan, X., Protchenko, O., Philpott, C.C. & Hamza, I. Topologically conserved residues direct heme transport in HRG-1-related proteins. *J Biol Chem* **287**, 4914-24 (2012).

76. O'Callaghan, K.M. *et al.* Heme-binding protein HRG-1 is induced by insulin-like growth factor I and associates with the vacuolar H+-ATPase to control endosomal pH and receptor trafficking. *J Biol Chem* **285**, 381-91 (2010).

77. Li, H., Mapolelo, D.T., Randeniya, S., Johnson, M.K. & Outten, C.E. Human glutaredoxin 3 forms [2Fe-2S]-bridged complexes with human BolA2. *Biochemistry* **51**, 1687-96 (2012).

78. Riley, L.G. *et al.* Phenotypic variability and identification of novel YARS2 mutations in YARS2 mitochondrial myopathy, lactic acidosis and sideroblastic anaemia. *Orphanet J Rare Dis* **8**, 193 (2013).

79. Sampson, N.D. & Hewitt, J.E. SF4 and SFRS14, two related putative splicing factors on human chromosome 19p13.11. *Gene* **305**, 91-100 (2003).

80. Cohen, D. *et al.* Pros and cons for C4d as a biomarker. *Kidney Int* **81**, 628-39 (2012).

81. Di Pietro, E., Wang, X.L. & MacKenzie, R.E. The expression of mitochondrial methylenetetrahydrofolate dehydrogenase-cyclohydrolase supports a role in rapid cell growth. *Biochim Biophys Acta* **1674**, 78-84 (2004).

82. Nilsson, R. *et al.* Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat Commun* **5**, 3128 (2014).

83. Smith, R.W., Blee, T.K. & Gray, N.K. Poly(A)-binding proteins are required for diverse biological processes in metazoans. *Biochem Soc Trans* **42**, 1229-37 (2014).

84. Koc, E.C. *et al.* Identification and characterization of CHCHD1, AURKAIP1, and CRIF1 as new members of the mammalian mitochondrial ribosome. *Front Physiol* **4**, 183 (2013).

85. Machado-Neto, J.A. *et al.* Stathmin 1 is involved in the highly proliferative phenotype of high-risk myelodysplastic syndromes and acute leukemia cells. *Leuk Res* **38**, 251-7 (2014).

86. Yang, J.J. *et al.* ZAK inhibits human lung cancer cell growth via ERK and JNK activation in an AP-1-dependent manner. *Cancer Sci* **101**, 1374-81 (2010).

87. Kamioka, M., Imamura, J., Komatsu, N., Daibata, M. & Sugiura, T. Testican 3 expression in adult T-cell leukemia. *Leuk Res* **33**, 913-8 (2009).

88. Zassadowski, F. *et al.* Auer rods and differentiation in acute promyelocytic leukemia. *British Journal of Haematology* **142**, 998-1000 (2014).

89. Rivas, V. *et al.* Developmental and tumoral vascularization is regulated by G protein-coupled receptor kinase 2. *J Clin Invest* **123**, 4714-30 (2013).

90. Legha, W. *et al.* stac1 and stac2 genes define discrete and distinct subsets of dorsal root ganglia neurons. *Gene Expr Patterns* **10**, 368-75 (2010).

91. Ishii, A., Zhang, B., Kaneko, S. & Hirose, S. Positive association between benign familial infantile convulsions and LGI4. *Brain Dev* **32**, 538-43 (2010).

92. Aslanyan, M.G. *et al.* Clinical and biological impact of TET2 mutations and expression in younger adult AML patients treated within the EORTC/GIMEMA AML-12 clinical trial. *Ann Hematol* **93**, 1401-12 (2014).

93. Nakajima, H. & Kunimoto, H. TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Sci* **105**, 1093-9 (2014).

94. Elias, H.K. *et al.* Stem cell origin of myelodysplastic syndromes. *Oncogene* **33**, 5139-50 (2014).

95. Pietras, E.M., Warr, M.R. & Passegue, E. Cell cycle regulation in hematopoietic stem cells. *J Cell Biol* **195**, 709-20 (2011).

96. DAVID Bioinformatics Resources. (2007). (last accessed on 15.01.2015)

97. O'Bleness, M.S. *et al.* Evolutionary history and genome organization of DUF1220 protein domains. *G3 (Bethesda)* **2**, 977-86 (2012).

98. Zhang, H.X. *et al.* Rig-I regulates NF-kappaB activity through binding to Nf-kappab1 3'-UTR mRNA. *Proc Natl Acad Sci U S A* **110**, 6459-64 (2013).

99. Kubota, T. *et al.* Amyloid beta (A4) precursor protein expression in human periodontitis-affected gingival tissues. *Arch Oral Biol* **59**, 586-94 (2014).

100. Oda, T. *et al.* ABT1-associated protein (ABTAP), a novel nuclear protein conserved from yeast to mammals, represses transcriptional activation by ABT1. *J Cell Biochem* **93**, 788-806 (2004).

101. Marzo, S., Galimberti, V. & Biggiogera, M. Unexpected distribution of KRIT1 inside the nucleus: new insight in a complex molecular pathway. *Eur J Histochem* **58**, 2358 (2014).

102. Wang, R., Nan, Y., Yu, Y. & Zhang, Y.J. Porcine reproductive and respiratory syndrome virus Nsp1beta inhibits interferon-activated JAK/STAT signal transduction by inducing karyopherin-alpha1 degradation. *J Virol* **87**, 5219-28 (2013).

103. Liao, W. *et al.* Silencing of DLGAP5 by siRNA significantly inhibits the proliferation and invasion of hepatocellular carcinoma cells. *PLoS One* **8**, e80789 (2013).

104. Gomez, C.R. *et al.* Prognostic value of discs large homolog 7 transcript levels in prostate cancer. *PLoS One* **8**, e82833 (2013).

105. Home - Gene - NCBI. (2014). (last accessed on 15.01.2015)

106. Barenz, F. *et al.* The centriolar satellite protein SSX2IP promotes centrosome maturation. *J Cell Biol* **202**, 81-95 (2013).

107. Hori, A., Ikebe, C., Tada, M. & Toda, T. Msd1/SSX2IP-dependent microtubule anchorage ensures spindle orientation and primary cilia formation. *EMBO Rep* **15**, 175-84 (2014).

108. Sjakste, T. *et al.* Juvenile idiopathic arthritis subtype- and sex-specific associations with genetic variants in the PSMA6/PSMC6/PSMA3 gene cluster. *Pediatr Neonatol* **55**, 393-403 (2014).

109. Kuhle, B. & Ficner, R. eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. *Embo j* **33**, 1177-91 (2014).

110. Patel, D. & Chaudhary, J. Increased expression of bHLH transcription factor E2A (TCF3) in prostate cancer promotes proliferation and confers resistance to doxorubicin induced apoptosis. *Biochem Biophys Res Commun* **422**, 146-51 (2012).

111. Slyper, M. *et al.* Control of breast cancer growth and initiation by the stem cell-associated transcription factor TCF3. *Cancer Res* **72**, 5613-24 (2012).

112. Shu, K.X. *et al.* Characterization of the human PAP1 gene and its homologue possible involvement in mouse embryonic development. *Colloids Surf B Biointerfaces* **52**, 22-30 (2006).

113. Chaikuad, A. *et al.* Structure of human aspartyl aminopeptidase complexed with substrate analogue: insight into catalytic mechanism, substrate specificity and M18 peptidase family. *BMC Struct Biol* **12**, 14 (2012).

114. Chen, Y., Farquhar, E.R., Chance, M.R., Palczewski, K. & Kiser, P.D. Insights into substrate specificity and metal activation of mammalian tetrahedral aspartyl aminopeptidase. *J Biol Chem* **287**, 13356-70 (2012).

115. Beser, A.R. *et al.* HER-2, TOP2A and chromosome 17 alterations in breast cancer. *Pathol Oncol Res* **13**, 180-5 (2007).

116. Dubrowinskaja, N. *et al.* Neurofilament Heavy polypeptide CpG island methylation associates with prognosis of renal cell carcinoma and prediction of antivascular endothelial growth factor therapy response. *Cancer Med* **3**, 300-9 (2014).

117. Schleicher, R.L. *et al.* Neurofilament heavy chain-like messenger RNA and protein are present in benign prostate and down-regulated in prostatic carcinoma. *Cancer Res* **57**, 3532-6 (1997).

118. Li, P. *et al.* Epigenetic silencing of miR-338-3p contributes to tumorigenicity in gastric cancer by targeting SSX2IP. *PLoS One* **8**, e66782 (2013).

119. Kumar, S. & Sharma, R.K. An improved method and cost effective strategy for soluble expression and purification of human N-myristoyltransferase 1 in E. coli. *Mol Cell Biochem* **392**, 175-86 (2014).

120. Muglia, M. *et al.* A novel KIF5A mutation in an Italian family marked by spastic paraparesis and congenital deafness. *J Neurol Sci* **343**, 218-20 (2014).

121. Donovan, J. & Copeland, P.R. Selenocysteine insertion sequence binding protein 2L is implicated as a novel post-transcriptional regulator of selenoprotein expression. *PLoS One* **7**, e35581 (2012).

122. Cai, M. *et al.* Atp6v1c1 may regulate filament actin arrangement in breast cancer cells. *PLoS One* **9**, e84833 (2014).

123. Feng, S. *et al.* Silencing of atp6v1c1 prevents breast cancer growth and bone metastasis. *Int J Biol Sci* **9**, 853-62 (2013).

124. Sotgia, F. *et al.* Mitochondria "fuel" breast cancer metabolism: fifteen markers of mitochondrial biogenesis label epithelial cancer cells, but are excluded from adjacent stromal cells. *Cell Cycle* **11**, 4390-401 (2012).

125. Fraisier, V. *et al.* C11ORF24 Is a Novel Type I Membrane Protein That Cycles between the Golgi Apparatus and the Plasma Membrane in Rab6-Positive Vesicles. *PLoS One* **8**, e82223 (2013).

126. Wu, Z. *et al.* Biglycan and decorin differentially regulate signaling in the fetal membranes. *Matrix Biol* **35**, 266-75 (2014).

127. Berendsen, A.D. *et al.* Biglycan modulates angiogenesis and bone formation during fracture healing. *Matrix Biol* **35**, 223-31 (2014).

128. Burger, M.C., De Wet, H. & Collins, M. The BGN and ACAN genes and carpal tunnel syndrome. *Gene* **551**, 160-6 (2014).

129. Hu, L. *et al.* Biglycan enhances gastric cancer invasion by activating FAK signaling pathway. *Oncotarget* **5**, 1885-96 (2014).

130. Chaudhuri, R., Mattera, R., Lindwasser, O.W., Robinson, M.S. & Bonifacino, J.S. A basic patch on alpha-adaptin is required for binding of human immunodeficiency virus type 1 Nef and cooperative assembly of a CD4-Nef-AP-2 complex. *J Virol* **83**, 2518-30 (2009).

131. Sharp, S.I. *et al.* The effect of clozapine on mRNA expression for genes encoding G protein-coupled receptors and the protein components of clathrin-mediated endocytosis. *Psychiatr Genet* **23**, 153-62 (2013).

132. Mancias, J.D. & Goldberg, J. The transport signal on Sec22 for packaging into COPII-coated vesicles is a conformational epitope. *Mol Cell* **26**, 403-14 (2007).

133. SEC22A SEC22 vesicle trafficking protein homolog A (S. cerevisiae) [Homo sapiens (human)] - Gene - NCBI. (Pubs, 2013). (last accessed on 15.01.2015)

134. Kim, Y.R. *et al.* Mutational and expressional analysis of RFC3, a clamp loader in DNA replication, in gastric and colorectal cancers. *Hum Pathol* **41**, 1431-7 (2010).

135. Lockwood, W.W. *et al.* Integrative genomics identified RFC3 as an amplified candidate oncogene in esophageal adenocarcinoma. *Clin Cancer Res* **18**, 1936-46 (2012).

136. Nigg, E.A., Cajanek, L. & Arquint, C. The centrosome duplication cycle in health and disease. *FEBS Lett* **588**, 2366-72 (2014).

137. Schmidt, K.N. *et al.* Cep164 mediates vesicular docking to the mother centriole during early steps of ciliogenesis. *J Cell Biol* **199**, 1083-101 (2012).

138. Yosten, G.L., Redlinger, L.J. & Samson, W.K. Evidence for an interaction of neuronostatin with the orphan G protein-coupled receptor, GPR107. *Am J Physiol Regul Integr Comp Physiol* **303**, R941-9 (2012).

139. Cheng, L. *et al.* A 4-gene panel as a marker at chromosome 8q in Asian gastric cancer patients. *Genomics* **102**, 323-30 (2013).

140. Lin, C.L., Leu, S., Lu, M.C. & Ouyang, P. Over-expression of SR-cyclophilin, an interaction partner of nuclear pinin, releases SR family splicing factors from nuclear speckles. *Biochem Biophys Res Commun* **321**, 638-47 (2004).

141. Djian, P., Easley, K. & Green, H. Targeted ablation of the murine involucrin gene. *J Cell Biol* **151**, 381-8 (2000).

142. Sevilla, L.M. *et al.* Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier. *J Cell Biol* **179**, 1599-612 (2007).

143. Glenn, K.A., Nelson, R.F., Wen, H.M., Mallinger, A.J. & Paulson, H.L. Diversity in tissue expression, substrate binding, and SCF complex formation for a lectin family of ubiquitin ligases. *J Biol Chem* **283**, 12717-29 (2008).

144. Bell, A., Bell, D., Weber, R.S. & El-Naggar, A.K. CpG island methylation profiling in human salivary gland adenoid cystic carcinoma. *Cancer* **117**, 2898-909 (2011).

145. Smith, N.L., Hankinson, J., Simpson, A., Denning, D.W. & Bowyer, P. Reduced expression of TLR3, TLR10 and TREM1 by human macrophages in CCPA, and novel associations of VEGFA, DENND1B and PLAT. *Clin Microbiol Infect* (2014).

146. Lango Allen, H. *et al.* Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC1I1 exonic enhancers of DLX5/6 expression in humans. *J Med Genet* **51**, 264-7 (2014).

147. Li, C.C. *et al.* GTP-binding-defective ARL4D alters mitochondrial morphology and membrane potential. *PLoS One* **7**, e43552 (2012).

148. Jones, A. *et al.* Role of DNA methylation and epigenetic silencing of HAND2 in endometrial cancer development. *PLoS Med* **10**, e1001551 (2013).

149. Line, A., Slucka, Z., Stengrevics, A., Li, G. & Rees, R.C. Altered splicing pattern of TACC1 mRNA in gastric cancer. *Cancer Genet Cytogenet* **139**, 78-83 (2002).

150. Xiang, J. *et al.* Efficient downregulation of ErbB-2 induces TACC1 upregulation in breast cancer cell lines. *Oncol Rep* **29**, 1517-23 (2013).

151. Singh, D. *et al.* Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337**, 1231-5 (2012).

152. Zhang, X.W. *et al.* Oncogenic role of the chromobox protein CBX7 in gastric cancer. *J Exp Clin Cancer Res* **29**, 114 (2010).

153. Guan, Z.P. *et al.* [Downregulation of chromobox protein homolog 7 expression in multiple human cancer tissues]. *Zhonghua Yu Fang Yi Xue Za Zhi* **45**, 597-600 (2011).

154. Zhang, L., Lubin, A., Chen, H., Sun, Z. & Gong, F. The deubiquitinating protein USP24 interacts with DDB2 and regulates DDB2 stability. *Cell Cycle* **11**, 4378-84 (2012).

155. Ishida, H. *et al.* Identification and characterization of novel isoforms of human DP-1: DP-1{alpha} regulates the transcriptional activity of E2F1 as well as cell cycle progression in a dominant-negative manner. *J Biol Chem* **280**, 24642-8 (2005).

156. Chen, T., Sun, Y., Ji, P., Kopetz, S. & Zhang, W. Topoisomerase IIalpha in chromosome instability and personalized cancer therapy. in *Oncogene* (2014).

157. Jarvinen, T.A. & Liu, E.T. Topoisomerase IIalpha gene (TOP2A) amplification and deletion in cancer--more common than anticipated. *Cytopathology* **14**, 309-13 (2003).

158. Varis, A. *et al.* Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res* **62**, 2625-9 (2002).

159. Simon, R. *et al.* [HER-2 and TOP 2A gene amplifications in urinary bladder carcinoma]. *Verh Dtsch Ges Pathol* **86**, 176-83 (2002).

160. Andoniadou, C.L. *et al.* HESX1- and TCF3-mediated repression of Wnt/beta-catenin targets is required for normal development of the anterior forebrain. *Development* **138**, 4931-42 (2011).

161. Logan, C.Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**, 781-810 (2004).

162. Liaw, H.H., Lin, C.C., Juan, H.F. & Huang, H.C. Differential microRNA regulation correlates with alternative polyadenylation pattern between breast cancer and normal cells. *PLoS One* **8**, e56958 (2013).

163. Mayr, C. & Bartel, D.P. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673-84 (2009).

164. An, J., Zhu, X., Wang, H. & Jin, X. A dynamic interplay between alternative polyadenylation and microRNA regulation: implications for cancer (Review). *Int J Oncol* **43**, 995-1001 (2013).

APPENDIX A

MICROARRAY DATASETS

Table A.1 GSE19429, GSE12662 and GSE30201 datasets.

GEO Series Number	GEO Accession Number	Disease State	Cell Type / Genetic Content
GSE12662	GSM317736	Healthy	$CD34^+$/Normal
GSE12662	GSM317739	Healthy	$CD34^+$/Normal
GSE12662	GSM317794	Healthy	$CD34^+$/Normal
GSE12662	GSM317933	Healthy	$CD34^+$/Normal
GSE12662	GSM317934	Healthy	$CD34^+$/Normal
GSE30201	GSM747521	Healthy	$CD34^+$/Normal
GSE30201	GSM747522	Healthy	$CD34^+$/Normal
GSE30201	GSM747523	Healthy	$CD34^+$/Normal
GSE30201	GSM747524	Healthy	$CD34^+$/Normal
GSE30201	GSM747525	Healthy	$CD34^+$/Normal
GSE30201	GSM747526	Healthy	$CD34^+$/Normal
GSE30201	GSM747527	Healthy	$CD34^+$/Normal
GSE30201	GSM747528	Healthy	$CD34^+$/Normal
GSE30201	GSM747529	Healthy	$CD34^+$/Normal
GSE30201	GSM747530	Healthy	$CD34^+$/Normal
GSE30201	GSM747531	Healthy	$CD34^+$/Normal
GSE19429	GSM483480	Healthy	$CD34^+$/Normal
GSE19429	GSM483481	Healthy	$CD34^+$/Normal
GSE19429	GSM483482	Healthy	$CD34^+$/Normal
GSE19429	GSM483483	Healthy	$CD34^+$/Normal
GSE19429	GSM483484	Healthy	$CD34^+$/Normal
GSE19429	GSM483485	Healthy	$CD34^+$/Normal
GSE19429	GSM483486	Healthy	$CD34^+$/Normal
GSE19429	GSM483487	Healthy	$CD34^+$/Normal
GSE19429	GSM483488	Healthy	$CD34^+$/Normal
GSE19429	GSM483489	Healthy	$CD34^+$/Normal
GSE19429	GSM483490	Healthy	$CD34^+$/Normal
GSE19429	GSM483491	Healthy	$CD34^+$/Normal
GSE19429	GSM483492	Healthy	$CD34^+$/Normal
GSE19429	GSM483493	Healthy	$CD34^+$/Normal
GSE19429	GSM483494	Healthy	$CD34^+$/Normal
GSE19429	GSM483495	Healthy	$CD34^+$/Normal
GSE19429	GSM483496	Healthy	$CD34^+$/Normal
GSE19429	GSM483313	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483328	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483333	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483336	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483337	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483339	MDS-RA	$CD34^+$/Normal

Table A.1 (continued)

GEO Series Number	GEO Accession Number	Disease State	Cell Type / Genetic Content
GSE19429	GSM483351	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483352	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483384	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483388	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483390	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483396	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483399	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483401	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483412	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483418	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483420	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483421	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483426	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483455	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483461	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483468	MDS-RA	$CD34^+$/Normal
GSE19429	GSM483297	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483298	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483299	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483314	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483316	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483321	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483326	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483335	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483340	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483355	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483357	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483360	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483382	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483387	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483392	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483394	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483395	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483398	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483423	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483429	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483430	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483435	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483438	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483441	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483451	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483452	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483467	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483469	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483472	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483475	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483479	MDS-RARS	$CD34^+$/Normal
GSE19429	GSM483323	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483338	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483353	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483363	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483364	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483368	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483373	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483374	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483379	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483381	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483389	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483404	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483410	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483416	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483419	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483427	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483445	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483459	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483465	MDS-RAEB1	$CD34^+$/Normal

Table A.1 (continued)

GEO Series Number	GEO Accession Number	Disease State	Cell Type / Genetic Content
GSE19429	GSM483470	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483478	MDS-RAEB1	$CD34^+$/Normal
GSE19429	GSM483343	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483347	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483348	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483349	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483350	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483362	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483265	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483369	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483370	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483375	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483378	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483408	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483424	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483437	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483439	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483440	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483446	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483449	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483460	MDS-RAEB2	$CD34^+$/Normal
GSE19429	GSM483471	MDS-RAEB2	$CD34^+$/Normal

Table A.2 GSE4290 dataset.

GEO Accession Number	Title	Source Name	Histopathological Diagnosis
GSM97800	HF0088_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97803	HF0120_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97804	HF0131_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97805	HF0137_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97807	HF0141_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97809	HF0151_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97811	HF0163_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97812	HF0171_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97816	HF0201_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97817	HF0201_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97820	HF0232_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97825	HF0295_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97827	HF0303_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97828	HF0312_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97833	HF0377_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97834	HF0383_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97840	HF0467_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97846	HF0512_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97848	HF0523_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97849	HF0526_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97850	HF0533_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97853	HF0593_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97855	HF0616_U133P2	Brain tissue from epilepsy patient	Non-tumor
GSM97794	HF0024_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97796	HF0031_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97797	HF0050_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97798	HF0066_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97801	HF0089_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97806	HF0138_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97808	HF0142_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97813	HF0180_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97814	HF0184_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97818	HF0212_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97819	HF0218_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97821	HF0244_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97829	HF0316_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97832	HF0350_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97839	HF0460_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97844	HF0505_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97847	HF0520_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97851	HF0543_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97852	HF0583_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97856	HF0627_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97859	HF0654_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97861	HF0702_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97863	HF0790_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97869	HF0850_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97870	HF0855_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97871	HF0894_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97877	HF0936_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97882	HF0963_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97885	HF0982_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97886	HF0986_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97887	HF0990_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97888	HF0992_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97889	HF0996_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97891	HF1057_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97892	HF1058_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97893	HF1077_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97894	HF1078_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97895	HF1097_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97896	HF1122_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97898	HF1137_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97903	HF1178_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97905	HF1186_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4

Table A.2 (continued)

GEO Accession Number	Title	Source Name	Histopathological Diagnosis
GSM97906	HF1191_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97908	HF1220_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97912	HF1242_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97914	HF1255_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97915	HF1262_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97917	HF1280_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97918	HF1286_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97919	HF1292_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97922	HF1318_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97924	HF1326_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97926	HF1338_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97930	HF1356_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97931	HF1357_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97935	HF1382_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97936	HF1397_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97938	HF1409_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97940	HF1458_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97942	HF1475_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97945	HF1490_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97946	HF1492_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97948	HF1494_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97950	HF1509_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97952	HF1517_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97953	HF1534_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97954	HF1538_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97955	HF1540_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97959	HF1585_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97961	HF1589_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97963	HF1608_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97965	HF1618_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97966	HF1628_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97967	HF1640_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97968	HF1640_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97969	HF1671_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4
GSM97971	HF1702_U133P2	Brain tissue from glioma patient	Glioblastoma, grade 4

Table A.3 GSE16011 dataset.

GEO Accession Number	Gender	Age	Surgery	Radiation	Chemotherapy	Live Status	Survival (years)
Control Samples							
GSM405200	-	-	-	-	-	-	-
GSM405209	-	-	-	-	-	-	-
GSM405332	-	-	-	-	-	-	-
GSM405336	-	-	-	-	-	-	-
GSM405357	-	-	-	-	-	-	-
GSM405358	-	-	-	-	-	-	-
GSM405359	-	-	-	-	-	-	-
GSM405360	-	-	-	-	-	-	-
Glioblastoma (grade IV)							
GSM405213	Male	44.09	Complete Resection	Yes	No	Dead	1.34
GSM405214	Male	37.84	Partial Resection	Yes	Adjuvant Temozolomide	Dead	1.50
GSM405215	Male	51.44	Partial Resection	Yes	No	Dead	2.30
GSM405216	Male	52.06	Complete Resection	Yes	No	Dead	3.28
GSM405217	Male	33.20	Partial Resection	Yes	Concomitant Temozolomide	Alive	6.77
GSM405218	Male	32.36	Partial Resection	Yes	Concomitant Temozolomide	Dead	0.64
GSM405219	Male	57.25	Partial Resection	Yes	Concomitant Temozolomide	Dead	0.58
GSM405220	Male	54.12	Complete Resection	Yes	Concomitant Temozolomide	Dead	1.27
GSM405221	Male	55.52	Complete Resection	Yes	No	Dead	0.69
GSM405222	Male	54.06	Complete Resection	Yes	No	Dead	1.30
GSM405223	Female	53.26	Complete Resection	Yes	No	Dead	1.92
GSM405224	Male	47.74	Complete Resection	Yes	No	Dead	0.48
GSM405228	Male	45.71	Open Biopsy	Yes	No	Dead	0.71
GSM405229	Female	54.18	Partial Resection	Yes	No	Dead	0.66
GSM405230	Female	63.33	Complete Resection	Yes	No	Dead	0.47
GSM405231	Female	62.96	Stereotactic Biopsy	Yes	No	Dead	1.26
GSM405232	Male	35.70	Partial Resection	Yes	No	Dead	0.98
GSM405233	Male	51.64	Stereotactic Biopsy	Yes	No	Dead	0.86
GSM405234	Female	57.68	Partial Resection	Yes	No	Dead	0.62
GSM405235	Male	71.09	Partial Resection	Yes	No	Dead	0.21
GSM405236	Male	52.20	Complete Resection	Yes	No	Dead	1.03
GSM405237	Male	51.17	Stereotactic Biopsy	Yes	No	Dead	0.41
GSM405238	Male	37.25	Stereotactic Biopsy	Yes	No	Dead	0.94
GSM405239	Female	73.65	Complete Resection	Yes	No	Dead	0.59
GSM405240	Male	33.09	Complete Resection	Yes	Adjuvant PCV	Alive	6.62
GSM405241	Male	55.98	Partial Resection	Yes	No	Dead	0.16
GSM405242	Male	30.33	Stereotactic Biopsy	Yes	No	Dead	0.18
GSM405243	Male	61.33	Partial Resection	Yes	No	Dead	0.88
GSM405244	Male	29.55	Partial Resection	Yes	No	Dead	2.05
GSM405245	Male	37.98	Partial Resection	Yes	No	Dead	1.40
GSM405246	Male	32.59	Partial Resection	Yes	No	Dead	6.31
GSM405247	Male	39.36	Partial Resection	Yes	Adjuvant Chemotherapeutic	Dead	1.59

Table A.3 (continued)

GEO Accession Number	Gender	Age	Surgery	Radiation	Chemotherapy	Live Status	Survival (years)
GSM405248	Male	55.20	Stereotactic Biopsy	Yes	No	Dead	0.84
GSM405249	Male	23.02	Stereotactic Biopsy	-	-	Dead	0.04
GSM405251	Female	26.17	Stereotactic Biopsy	Yes	No	Dead	3.21
GSM405252	Male	68.04	Complete Resection	Yes	No	Dead	0.37
GSM405253	Male	34.84	Complete Resection	Yes	No	Alive	12.56
GSM405254	Male	40.58	Partial Resection	Yes	No	Dead	0.59
GSM405255	Male	47.93	Partial Resection	Yes	No	Dead	1.08
GSM405260	Male	57.58	Partial Resection	Yes	No	Dead	0.82
GSM405262	Male	43.26	Complete Resection	Yes	No	Dead	2.89
GSM405263	Male	61.74	Partial Resection	Yes	Concomitant Chemotherapeutic	Dead	1.55
GSM405264	Male	36.50	Partial Resection	Yes	No	Dead	0.32
GSM405266	Male	51.62	Partial Resection	-	-	Dead	0.25
GSM405267	Female	53.30	Stereotactic Biopsy	Yes	No	Dead	0.65
GSM405268	Male	48.04	Partial Resection	Yes	No	Dead	0.64
GSM405269	Male	54.59	Partial Resection	-	-	Dead	0.03
GSM405270	Female	73.11	Partial Resection	Yes	No	Dead	0.58
GSM405271	Male	69.89	Complete Resection	Yes	No	Dead	0.30
GSM405274	Female	63.08	Complete Resection	Yes	No	Dead	1.29
GSM405275	Female	60.39	Complete Resection	Yes	No	Dead	3.46
GSM405276	Male	51.85	Partial Resection	Yes	No	Dead	0.28
GSM405278	Male	58.23	Complete Resection	Yes	No	Dead	0.73
GSM405280	Female	30.43	Complete Resection	Yes	No	Dead	2.46
GSM405282	Male	51.92	Stereotactic Biopsy	-	-	Dead	0.12
GSM405290	Female	45.50	Stereotactic Biopsy	Yes	No	Dead	1.16
GSM405292	Male	65.52	Partial Resection	Yes	No	Dead	1.11
GSM405293	Male	63.73	Partial Resection	Yes	No	Dead	0.88
GSM405294	Male	56.41	Complete Resection	Yes	No	Dead	0.80
GSM405296	Female	41.09	Stereotactic Biopsy	Yes	No	Dead	0.29
GSM405297	Female	67.13	Complete Resection	Yes	No	Dead	0.72
GSM405299	Female	54.94	Complete Resection	Yes	No	Dead	1.75
GSM405301	Male	65.35	Complete Resection	-	-	Dead	0.30
GSM405302	Male	41.39	Partial Resection	Yes	No	Dead	0.74
GSM405303	Male	56.64	Partial Resection	Yes	No	Dead	0.55
GSM405304	Male	66.39	Stereotactic Biopsy	-	-	Dead	0.56
GSM405305	Male	52.69	Partial Resection	Yes	No	Dead	1.06
GSM405307	Male	72.22	Complete Resection	-	-	Dead	0.33
GSM405308	Male	47.40	Partial Resection	Yes	No	-	3.10
GSM405309	Female	54.60	Partial Resection	-	-	Dead	0.26
GSM405312	Male	61.31	Partial Resection	Yes	No	Dead	1.02
GSM405313	Male	32.50	Complete Resection	Yes	No	Dead	3.31

Table A.3 (continued)

GEO Accession Number	Gender	Age	Surgery	Radiation	Chemotherapy	Live Status	Survival (years)
GSM405314	Male	54.37	Complete Resection	Yes	No	Dead	0.65
GSM405315	Male	64.61	Partial Resection	Yes	No	Dead	0.35
GSM405317	Female	62.36	Partial Resection	Yes	No	Dead	0.45
GSM405320	Male	70.28	Partial Resection	Yes	No	Dead	0.60
GSM405322	Male	62.28	Partial Resection	Yes	No	Dead	1.14
GSM405323	Female	58.78	Complete Resection	Yes	No	Dead	0.62
GSM405324	Male	14.38	Complete Resection	Yes	No	Dead	0.67
GSM405326	Male	74.56	Partial Resection	-	-	Dead	0.27
GSM405328	Male	67.60	Complete Resection	Yes	No	Dead	2.26
GSM405330	Female	33.12	Partial Resection	Yes	No	Dead	0.71
GSM405337	Female	15.02	Partial Resection	-	Upfront Temozolomide	Dead	0.28
GSM405339	Female	78.08	Partial Resection	-	-	-	-
GSM405340	Female	52.88	Open Biopsy	Yes	No	-	5.56
GSM405343	Male	66.80	Partial Resection	-	-	-	-
GSM405345	Male	45.63	Partial Resection	Yes	No	Dead	0.89
GSM405349	Female	24.42	Complete Resection	Yes	No	Dead	2.41
GSM405350	Female	40.74	Complete Resection	Yes	No	Dead	0.61
GSM405351	Female	42.70	Partial Resection	Yes	No	Dead	1.22
GSM405352	Female	69.95	Partial Resection	Yes	-	Dead	0.40
GSM405353	Male	55.39	Partial Resection	Yes	No	Dead	0.70
GSM405356	Female	43.21	Partial Resection	Yes	No	Dead	0.18
GSM405362	Male	38.11	Stereotactic Biopsy	Yes	No	Dead	1.06
GSM405363	Female	48.84	Complete Resection	-	-	Dead	9.79
GSM405365	Male	64.29	Complete Resection	-	-	Dead	2.66
GSM405367	Male	46.83	Partial Resection	Yes	No	Dead	2.62
GSM405368	Female	32.14	Complete Resection	Yes	No	Dead	1.81
GSM405369	Female	50.34	Partial Resection	Yes	No	Dead	4.13
GSM405370	Male	46.52	Complete Resection	Yes	No	Dead	1.61
GSM405371	Female	53.26	Complete Resection	Yes	No	Dead	1.92
GSM405372	Male	37.12	Partial Resection	Yes	No	Dead	3.32
GSM405373	Male	58.58	Partial Resection	Yes	No	Dead	1.21
GSM405374	Male	58.58	Partial Resection	Yes	No	Dead	1.21
GSM405375	Female	67.03	Partial Resection	-	-	Dead	0.06
GSM405376	Male	52.98	Partial Resection	Yes	No	Dead	1.85
GSM405379	Male	60.28	Partial Resection	-	-	Dead	0.05
GSM405384	Female	70.30	Partial Resection	-	-	-	0.02
GSM405385	Male	34.78	Complete Resection	Yes	No	-	1.26
GSM405389	Female	52.88	Open Biopsy	Yes	No	-	5.56
GSM405391	Female	55.55	Partial Resection	Yes	No	Dead	1.05
GSM405392	Male	48.35	Partial Resection	Yes	No	Dead	0.47
GSM405393	Male	70.67	Partial Resection	-	-	Dead	0.08
GSM405396	Male	77.31	Partial Resection	-	-	Dead	0.02
GSM405397	Male	60.36	Partial Resection	-	-	Dead	0.35
GSM405398	Male	56.85	-	-	-	Dead	-
GSM405405	Male	71.09	Partial Resection	Yes	No	Dead	0.61
GSM405412	Male	55.71	Partial Resection	-	-	Dead	0.65
GSM405415	Male	67.48	Complete Resection	-	-	Dead	0.50
GSM405416	Male	64.26	Open Biopsy	-	-	Dead	0.34
GSM405417	Male	77.31	Partial Resection	-	-	Dead	0.02

Table A.3 (continued)

GEO Accession Number	Gender	Age	Surgery	Radiation	Chemotherapy	Live Status	Survival (years)
GSM405418	Male	46.61	Stereotactic Biopsy	Yes	No	Dead	5.17
GSM405419	Male	36.27	Partial Resection	Yes	No	Dead	2.93
GSM405422	Female	70.67	Partial Resection	Yes	No	Dead	0.91
GSM405426	Male	67.10	Complete Resection	-	-	Dead	0.05
GSM405427	Female	50.83	Stereotactic Biopsy	Yes	No	Dead	1.53
GSM405428	Female	70.51	Partial Resection	Yes	No	Dead	0.79
GSM405430	Male	61.10	Complete Resection	Yes	No	Dead	0.35
GSM405431	Male	80.65	Partial Resection	Yes	No	Dead	0.92
GSM405432	Male	60.78	Partial Resection	Yes	No	Dead	0.55
GSM405434	Male	67.01	Complete Resection	Yes	No	Dead	0.24
GSM405436	Male	79.06	Partial Resection	Yes	No	Dead	0.48
GSM405438	Female	42.67	Partial Resection	Yes	Adjuvant Temozolomide	Dead	2.30
GSM405440	Male	69.88	Complete Resection	Yes	No	Dead	0.53
GSM405442	Female	63.61	Open Biopsy	Yes	Concomitant Temozolomide	Dead	0.30
GSM405443	Male	56.62	Stereotactic Biopsy	-	-	Dead	0.98
GSM405446	Male	68.18	Partial Resection	Yes	No	Dead	0.73
GSM405447	Male	59.03	Complete Resection	Yes	No	Dead	2.79
GSM405448	Female	58.71	Partial Resection	Yes	No	Dead	0.48
GSM405452	Female	71.02	Partial Resection	-	-	Dead	0.35
GSM405453	Male	70.23	Complete Resection	-	-	Dead	0.21
GSM405454	Male	22.52	Complete Resection	Yes	-	Dead	3.32
GSM405455	Female	55.49	Partial Resection	-	-	Dead	0.23
GSM405456	Male	52.50	-	-	-	-	-
GSM405458	Male	68.46	Complete Resection	Yes	No	Dead	1.58
GSM405459	Female	64.28	-	-	-	Dead	1.14
GSM405461	Female	49.14	Partial Resection	Yes	No	Dead	0.76
GSM405463	Female	65.53	Partial Resection	-	-	Dead	2.22
GSM405464	Male	54.72	Partial Resection	Yes	No	Dead	0.56
GSM405465	Male	69.38	Stereotactic Biopsy	-	-	Dead	0.63
GSM405466	Male	67.11	Stereotactic Biopsy	-	-	Dead	0.28
GSM405470	Male	31.72	Partial Resection	Yes	No	Dead	1.92
GSM405471	Female	78.12	Stereotactic Biopsy	-	-	Dead	0.15
GSM405472	Female	62.11	Complete Resection	-	-	Dead	0.34
GSM405473	Male	47.29	Partial Resection	Yes	No	Dead	1.12
GSM405474	Female	61.28	Partial Resection	-	-	Dead	0.29
GSM405475	Female	33.74	Complete Resection	Yes	No	Dead	1.05
GSM405477	Male	73.64	Stereotactic Biopsy	-	-	Dead	0.11
GSM405479	Male	63.30	Partial Resection	Yes	No	Dead	0.38

Table A.4 GSE4271 dataset.

GEO Accession Number	ID	WHO Grade	Age	Sex	Survival (weeks)
Primary Tumor Samples					
GSM96950	3744	IV with necrosis	60	Male	131
GSM96952	3746	IV with necrosis	43	Male	313
GSM96954	3748	IV with necrosis	45	Male	70
GSM96963	3758	IV with necrosis	34	Male	203
GSM96965	3774	IV with necrosis	48	Male	97
GSM96972	3781	IV with necrosis	43	Female	210
GSM96994	3906	IV with necrosis	32	Female	39
GSM96997	4094	IV with necrosis	32	Female	51
GSM97000	4097	IV with necrosis	57	Male	62
GSM97002	4102	IV with necrosis	68	Female	91
GSM97004	4122	IV with necrosis	72	Female	65
GSM97009	4127	IV with necrosis	72	Male	57
GSM97010	4128	IV with necrosis	57	Male	70
GSM97020	4755	III	25	Male	175
GSM97023	4759	III	43	Male	47
GSM97026	4767	III	44	Male	174
GSM97033	4778	III	23	Male	115
GSM97034	4779	III	34	Male	108
GSM97035	4780	III	32	Female	244
GSM97037	4782	IV without necrosis	49	Female	73
GSM97038	4783	III	45	Male	46
GSM97040	4785	IV without necrosis	54	Male	53
GSM97048	9907	IV with necrosis	25	Male	146
Recurrent Tumor Samples					
GSM96968	3777	IV	60	Male	131
GSM96975	3785	IV	43	Male	313
GSM96971	3780	IV	45	Male	70
GSM96956	3751	IV	34	Male	203
GSM96957	3752	IV	48	Male	97
GSM96960	3755	IV	43	Female	210
GSM96986	3840	IV	32	Female	39
GSM96998	4095	IV	32	Female	51
GSM97001	4098	IV	57	Male	62
GSM97003	4103	IV	68	Female	91
GSM97006	4124	IV	72	Female	65
GSM97005	4123	IV	72	Male	57
GSM96999	4096	IV	57	Male	70
GSM97043	4790	III	25	Male	175
GSM97013	4746	III	43	Male	47
GSM97021	4756	IV	44	Male	174
GSM97032	4777	IV	23	Male	115
GSM97029	4773	III	34	Male	108
GSM97024	4761	IV	32	Female	244
GSM97044	4791	IV	49	Female	73
GSM97031	4776	IV	45	Male	46
GSM97028	4771	IV	54	Male	53
GSM97049	9938	IV	25	Male	146

Table A.5 GSE29272 dataset - Cardia Patient Samples

GEO Accession Number - Adjacent Normal	GEO Accession Number - Tumor Sample	ID	Age	Sex	FH of UGI Cancer*	Cause of Death	Survival (days)
GSM723608	GSM723609	TYC0001	52	F	N	Mets**	172
GSM723610	GSM723611	TYC0002	37	F	Y	Mets	765
GSM723612	GSM723613	TYC0003	61	M	Y	Cancer	78
GSM723614	GSM723615	TYC0004	57	F	N	Cancer	521
GSM723616	GSM723617	TYC0005	69	F	Y	Cancer	1068
GSM723618	GSM723619	TYC0006	71	M	N	Mets	158
GSM723620	GSM723621	TYC0007	73	M	Y	Unknown	Unknown
GSM723622	GSM723623	TYC0008	37	F	N	Unknown	Unknown
GSM723624	GSM723625	TYC0009	62	F	Y	Cancer	398
GSM723626	GSM723627	TYC0010	66	M	N	Mets	895
GSM723628	GSM723629	TYC0011	61	M	N	Cancer	257
GSM723630	GSM723631	TYC0012	35	M	N	-	2392
GSM723632	GSM723633	TYC0013	57	M	N	Mets	253
GSM723634	GSM723635	TYC0014	60	M	N	Mets	1716
GSM723636	GSM723637	TYC0015	58	M	Y	Mets	608
GSM723638	GSM723639	TYC0016	58	M	N	Mets	136
GSM723640	GSM723641	TYC0017	62	M	N	Mets	167
GSM723642	GSM723643	TYC0018	59	M	N	Mets	557
GSM723644	GSM723645	TYC0019	69	M	N	Mets	206
GSM723646	GSM723647	TYC0020	68	M	N	Cancer	776
GSM723648	GSM723649	TYC0021	67	M	N	Cancer	378
GSM723650	GSM723651	TYC0022	61	M	Y	Unknown	Unknown
GSM723652	GSM723653	TYC0023	62	M	Y	Mets	364
GSM723654	GSM723655	TYC0024	69	M	N	Cancer	777
GSM723656	GSM723657	TYC0025	60	F	N	-	2035
GSM723658	GSM723659	TYC0026	58	M	N	Mets	76
GSM723660	GSM723661	TYC0027	58	M	N	-	2066
GSM723662	GSM723663	TYC0028	35	F	N	-	2736
GSM723664	GSM723665	TYC0029	62	M	N	Other	1363
GSM723666	GSM723667	TYC0030	66	M	Y	Other	1663
GSM723668	GSM723669	TYC0031	65	M	N	Unknown	Unknown
GSM723670	GSM723671	TYC0032	67	M	N	Mets	1204
GSM723672	GSM723673	TYC0033	67	M	N	Mets	559
GSM723674	GSM723675	TYC0034	48	M	Y	Mets	1328
GSM723676	GSM723677	TYC0035	52	F	Y	-	2309
GSM723678	GSM723679	TYC0036	53	M	N	Mets	215
GSM723680	GSM723681	TYC0037	57	M	N	Cancer	179
GSM723682	GSM723683	TYC0038	59	M	N	Mets	376
GSM723684	GSM723685	TYC0039	68	M	N	Cancer	525
GSM723686	GSM723687	TYC0040	46	M	N	Mets	417
GSM723688	GSM723689	TYC0041	64	M	N	Cancer	372
GSM723690	GSM723691	TYC0042	55	M	N	-	1742
GSM723692	GSM723693	TYC0043	62	F	N	Mets	647
GSM723694	GSM723695	TYC0044	66	F	Y	Mets	226
GSM723696	GSM723697	TYC0045	69	F	N	Mets	667
GSM723698	GSM723699	TYC0046	23	F	N	Mets	306
GSM723700	GSM723701	TYC0047	59	F	N	Unknown	Unknown
GSM723702	GSM723703	TYC0048	62	M	Y	Mets	872
GSM723704	GSM723705	TYC0049	54	F	N	Mets	396
GSM723706	GSM723707	TYC0050	41	M	Y	-	2227
GSM723708	GSM723709	TYC0051	65	M	N	Mets	97
GSM723710	GSM723711	TYC0052	58	M	N	Cancer	170
GSM723712	GSM723713	TYC0053	68	M	N	Mets	836
GSM723714	GSM723715	TYC0054	61	F	N	Unknown	Unknown
GSM723716	GSM723717	TYC0055	65	F	N	Mets	977
GSM723718	GSM723719	TYC0056	51	M	N	-	2092
GSM723720	GSM723721	TYC0057	63	M	N	Mets	134
GSM723722	GSM723723	TYC0058	66	M	Y	-	2187
GSM723724	GSM723725	TYC0059	68	M	N	Other	1480
GSM723726	GSM723727	TYC0060	55	M	N	-	1833
GSM723728	GSM723729	TYC0061	59	M	N	Unknown	Unknown
GSM723730	GSM723731	TYC0062	50	M	Y	-	1033

***FH of UGI Cancer:** Family History of Upper Gastrointestinal Cancer

****Mets:** Metastasis

Table A.6 GSE29272 dataset - Noncardia Patient Samples

GEO Accession Number - Adjacent Normal	GEO Accession Number - Tumor Sample	ID	Age	Sex	FH of UGI Cancer*	Cause of Death	Survival (days)
GSM723464	GSM723465	TYB0001	56	M	N	Mets**	881
GSM723466	GSM723467	TYB0002	49	M	N	-	2630
GSM723468	GSM723469	TYB0003	57	M	N	Mets	879
GSM723470	GSM723471	TYB0004	65	M	N	Mets	996
GSM723472	GSM723473	TYB0005	54	M	Y	Cancer	548
GSM723474	GSM723475	TYB0006	61	M	N	Mets	1078
GSM723476	GSM723477	TYB0007	55	M	N	Mets	608
GSM723478	GSM723479	TYB0008	66	F	N	Mets	894
GSM723480	GSM723481	TYB0009	59	F	N	Mets	189
GSM723482	GSM723483	TYB0010	60	M	N	Mets	798
GSM723484	GSM723485	TYB0011	52	M	N	Mets	586
GSM723486	GSM723487	TYB0012	62	M	Y	Mets	841
GSM723488	GSM723489	TYB0013	70	M	Y	Mets	462
GSM723490	GSM723491	TYB0014	64	M	N	Mets	161
GSM723492	GSM723493	TYB0015	35	F	Y	Mets	1066
GSM723494	GSM723495	TYB0016	58	M	Y	-	2790
GSM723496	GSM723497	TYB0017	64	M	N	-	2550
GSM723498	GSM723499	TYB0018	63	M	N	Mets	1330
GSM723500	GSM723501	TYB0019	59	M	N	-	2728
GSM723502	GSM723503	TYB0020	27	F	N	Mets	490
GSM723504	GSM723505	TYB0021	42	M	N	Mets	708
GSM723506	GSM723507	TYB0022	28	F	N	Mets	153
GSM723508	GSM723509	TYB0023	59	F	N	Mets	207
GSM723510	GSM723511	TYB0024	65	M	N	-	2466
GSM723512	GSM723513	TYB0025	28	F	Y	Mets	215
GSM723514	GSM723515	TYB0026	38	M	N	-	2810
GSM723516	GSM723517	TYB0027	61	M	N	Mets	869
GSM723518	GSM723519	TYB0028	57	F	Y	Unknown	Unknown
GSM723520	GSM723521	TYB0029	58	M	Y	Mets	240
GSM723522	GSM723523	TYB0030	60	M	N	Mets	279
GSM723524	GSM723525	TYB0031	67	M	Y	Cancer	821
GSM723526	GSM723527	TYB0032	59	M	Y	-	2081
GSM723528	GSM723529	TYB0033	57	M	N	-	2421
GSM723530	GSM723531	TYB0034	51	M	Y	Mets	288
GSM723532	GSM723533	TYB0035	43	F	N	Mets	93
GSM723534	GSM723535	TYB0036	55	M	N	Mets	364
GSM723536	GSM723537	TYB0037	67	M	Y	Mets	120
GSM723538	GSM723539	TYB0038	44	F	N	Mets	1053
GSM723540	GSM723541	TYB0039	57	M	N	Mets	169
GSM723542	GSM723543	TYB0040	47	M	N	Mets	243
GSM723544	GSM723545	TYB0041	39	M	N	Mets	881
GSM723546	GSM723547	TYB0042	63	M	N	-	2630
GSM723548	GSM723549	TYB0043	57	M	N	Mets	879
GSM723550	GSM723551	TYB0044	49	M	N	Mets	996
GSM723552	GSM723553	TYB0045	60	M	N	-	1855
GSM723554	GSM723555	TYB0046	59	M	N	Cancer	137
GSM723556	GSM723557	TYB0047	44	F	N	Mets	631
GSM723558	GSM723559	TYB0048	68	M	N	Mets	314
GSM723560	GSM723561	TYB0049	50	M	Y	-	1870
GSM723562	GSM723563	TYB0050	63	M	Y	Mets	379
GSM723564	GSM723565	TYB0051	55	M	N	-	2478
GSM723566	GSM723567	TYB0052	62	M	N	Mets	325
GSM723568	GSM723569	TYB0053	47	M	N	-	2502
GSM723570	GSM723571	TYB0054	43	M	N	Mets	829
GSM723572	GSM723573	TYB0055	57	M	Y	Mets	436
GSM723574	GSM723575	TYB0056	66	M	N	Mets	1191
GSM723576	GSM723577	TYB0057	60	M	N	Mets	248
GSM723578	GSM723579	TYB0058	64	M	N	-	1644
GSM723580	GSM723581	TYB0059	67	M	N	Mets	860
GSM723582	GSM723583	TYB0060	33	M	N	-	1863
GSM723584	GSM723585	TYB0061	61	M	N	-	1846
GSM723586	GSM723587	TYB0062	60	M	N	Mets	1762
GSM723588	GSM723589	TYB0063	63	M	N	-	1968
GSM723590	GSM723591	TYB0064	61	M	N	Mets	138

Table A.6 (continued)

GEO Accession Number - Adjacent Normal	GEO Accession Number - Tumor Sample	ID	Age	Sex	FH of UGI Cancer	Cause of Death	Survival (days)
GSM723592	GSM723593	TYB0065	50	M	N	Mets	364
GSM723594	GSM723595	TYB0066	54	F	N	Mets	892
GSM723596	GSM723597	TYB0067	59	M	N	Mets	848
GSM723598	GSM723599	TYB0068	54	M	N	-	2014
GSM723600	GSM723601	TYB0069	60	F	N	Mets	1462
GSM723602	GSM723603	TYB0070	42	F	N	-	1891
GSM723604	GSM723605	TYB0071	45	M	Y	Mets	350
GSM723606	GSM723607	TYB0072	68	M	N	Cancer	95

***FH of UGI Cancer:** Family History of Upper Gastrointestinal Cancer

****Mets:** Metastasis

APPENDIX B

GENES UNDERGOING SIGNIFICANT ALTERATIONS IN APA IDENTIFIED BY

APADETECT AND SAM

Table B.1 Significantly Altered Genes in RA Subtype of MDS

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
SCARB2	201647_s_at	Hs.349656.1.10	0.3630364	4.724485	0.3998154	0.08462624	0
RTN1	203485_at	Hs.368626.1.2	-0.43096477	4.3088875	0.43101385	0.10002904	0
SCARB2	201647_s_at	Hs.349656.1.9	0.3613597	4.280161	0.32273704	0.07540301	0
CFLAR	210564_x_at	Hs.390736.1.25	0.08349025	3.6978276	0.42639694	0.11531012	0
CFLAR	210564_x_at	Hs.390736.1.24	0.081868425	3.6978276	0.42639694	0.11531012	0
APIP	218698_at	Hs.447794.1.3	-0.74646103	3.6726544	0.570842	0.15543038	0
SNX11	53912_at	Hs.15827.1.19	-0.1933375	3.6632519	0.5761843	0.15728764	0
FLII	212024_x_at	Hs.513984.1.1	-0.22434756	3.5878997	0.5863389	0.1634212	0
POLR2C	216282_x_at	Hs.79402.1.17	-0.2838987	3.567567	0.5982778	0.1676991	0
YARS2	218470_at	Hs.505231.1.4	-0.6148896	3.5598164	0.48803663	0.13709602	0
C9orf82	231995_at	Hs.178357.1.3	1.0596745	3.5151966	0.3243232	0.09226318	0
POLR2C	216282_x_at	Hs.79402.1.18	-0.28218842	3.394393	0.56291413	0.16583647	0
LARS	222428_s_at	Hs.432674.1.6	0.5470252	3.3771174	0.49452245	0.14643331	0
GTPBP8	223486_at	Hs.127496.1.12	0.24968638	3.3682146	0.31040144	0.092156075	0
ACTL6A	202666_s_at	Hs.435326.1.30	0.28930545	3.3673658	0.44375086	0.13177982	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.8216283	3.35633	0.68565667	0.20428762	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.82463837	3.35633	0.68565667	0.20428762	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.81876945	3.35633	0.68565667	0.20428762	0
SLC46A3	214719_at	Hs.117167.1.1	-0.5168302	3.348591	0.4705571	0.14052391	0
CYP20A1	219565_at	Hs.446065.1.25	0.09179011	3.3354042	0.37524778	0.11250444	0
CYP20A1	219565_at	Hs.446065.1.24	0.089934796	3.3354042	0.37524778	0.11250444	0
ZKSCAN5	203731_s_at	Hs.110839.1.21	0.7882257	3.2916012	0.2397623	0.07284063	0
SNX13	1553148_a_at	Hs.487648.1.1	0.73319125	3.289141	0.47951365	0.1457869	0
TPI1	210050_at	Hs.524219.1.15	-0.6368717	3.2811816	0.5208025	0.15872407	0
SPOCK3	235342_at	Hs.481133.1.3	0.4216568	3.2578745	0.33837143	0.103862636	0
VTI1A	235034_at	Hs.194554.1.28	-0.8548226	3.254932	0.24199346	0.074346706	0
POLR2C	216282_x_at	Hs.79402.1.16	-0.28556168	3.24197	0.5406604	0.16676909	0
NNT	202783_at	Hs.482043.1.62	0.46133444	3.2275317	0.45078003	0.13966711	0
C19orf42	219097_x_at	Hs.356467.1.24	-0.08052494	3.1443093	0.32664984	0.10388604	0
CFLAR	210564_x_at	Hs.390736.1.14	0.080455855	3.1014125	0.3531838	0.11387837	0
NIP30	224248_x_at	Hs.396740.1.5	-0.2902	3.061546	0.39468122	0.12891565	0
LLPH	224446_at	Hs.504820.1.6	-0.5729675	3.0444415	0.39047784	0.12825927	0
PHKB	202739_s_at	Hs.78060.1.71	-0.2967425	3.0300038	0.2435813	0.08038977	0
SNX3	213545_x_at	Hs.12102.1.3	0.6708496	3.0001352	0.37356603	0.1245164	0
LYPLA2	215568_x_at	Hs.533479.1.18	-1.800125	2.998588	0.38710034	0.1290942	0
C22orf25	235396_at	Hs.474233.1.27	0.18739784	2.9946198	0.27786434	0.09278785	0

Table B.1 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
RAB3C	242328_at	Hs.482173.1.9	0.4647451	2.992572	0.30357632	0.101443276	0
BCLAF1	214499_s_at	Hs.486542.1.16	0.6943144	2.948171	0.45731544	0.15511836	0
CDC6	203968_s_at	Hs.405958.1.22	-0.21190418	2.941972	0.29815072	0.10134383	0
CREB1	204314_s_at	Hs.516646.1.28	0.09440483	2.9409428	0.28608286	0.097275905	0
ECHDC1	219974_x_at	Hs.486410.1.6	0.68453234	2.9272568	0.3123303	0.10669727	0
MYL4	217274_x_at	Hs.463300.1.13	-0.19818823	2.911431	0.24460608	0.08401576	0
INCENP	219769_at	Hs.142179.1.21	-0.7319496	2.9049668	0.32523495	0.11195823	0
SNX3	200067_x_at	Hs.12102.1.3	0.6686333	2.8901167	0.27877593	0.09645837	0
ADH5	208847_s_at	Hs.78989.1.10	0.38126323	2.882812	0.27309275	0.094731376	0
SF4	209547_s_at	Hs.515274.1.3	-0.07751301	2.8317995	0.39274824	0.13869211	0
C9orf85	238579_at	Hs.534190.1.18	1.0776174	2.8222058	0.2878056	0.10197895	0
GUF1	218884_s_at	Hs.546419.1.31	0.33995315	2.8082361	0.2500005	0.08902404	0
ADRBK1	38447_at	Hs.83636.1.51	-0.72696304	2.8065517	0.32102716	0.11438491	0
C12orf4	218374_s_at	Hs.302977.1.4	-0.6391299	2.7754858	0.20669518	0.07447171	0
EBNA1BP2	201323_at	Hs.346868.1.2	-1.4566485	2.7448142	0.40134794	0.14622045	0
PCMT1	208857_s_at	Hs.279257.1.17	0.69889706	2.7416668	0.27024287	0.09856883	0
NSDHL	215093_at	Hs.57698.1.21	2.1568406	2.7394354	0.36920267	0.13477327	0
VPS4B	218171_at	Hs.126550.1.2	-0.12158307	2.7388418	0.35537028	0.12975203	0
SET	213047_x_at	Hs.436687.1.19	1.1975467	2.7388349	0.2657218	0.097020015	0
METTL6	1553689_s_at	Hs.149487.1.8	0.22311828	2.7298343	0.14846084	0.05438456	0
CHEK1	205394_at	Hs.24529.1.27	-0.64305615	2.7144742	0.23023082	0.084815994	0
SLC48A1	48106_at	Hs.438867.1.13	-0.6090024	2.7142437	0.34625077	0.12756805	0
DNPEP	38703_at	Hs.258551.1.5	0.09889084	2.6813393	0.22310364	0.08320605	0
NIP30	223406_x_at	Hs.396740.1.6	-0.28717318	2.6694353	0.25111938	0.094072096	0
TRIB2	202479_s_at	Hs.467751.1.23	-0.026225684	2.6544225	0.25589916	0.096404836	0
UNKNOWN	235386_at	Hs.400256.1.1	0.31206948	2.654134	0.20624658	0.07770768	0
UNKNOWN	235386_at	Hs.400256.1.2	0.31364834	2.654134	0.20624658	0.07770768	0
MBNL3	219814_at	Hs.105134.1.8	1.8747596	2.6530895	0.34857106	0.13138308	0
JRKL	206734_at	Hs.105940.1.2	-0.69110334	2.6451335	0.24870968	0.09402538	0
C10orf84	218390_s_at	Hs.372309.1.3	-0.84729624	2.6440256	0.2517535	0.09521599	0
ATP7A	205198_s_at	Hs.496414.1.21	1.4186273	2.614813	0.3115024	0.11912989	0
KLHL20	210634_at	Hs.495035.1.12	-1.1293061	2.6147363	0.22156543	0.0847372	0
YARS2	218470_at	Hs.505231.1.6	-0.6125993	2.5991352	0.33442163	0.1286665	0
LOC440131	233804_at	Hs.132371.1.1	-0.5150005	2.5857415	0.13938911	0.053906824	0
LDB1	35160_at	Hs.454418.1.3	-0.8576225	2.5706327	0.30629957	0.11915338	0
ACOX3	204241_at	Hs.479122.1.6	0.3168265	2.5702617	0.23980325	0.09329916	0
KLHL24	242088_at	Hs.407709.1.19	0.29200184	2.5636704	0.32289618	0.12595074	0
ENOX2	234003_at	Hs.171458.1.30	1.8280462	2.5561192	0.24489242	0.09580634	0
SDF4	217855_x_at	Hs.42806.1.2	-2.3005052	2.5479841	0.30151373	0.11833423	0
BCLAF1	201083_s_at	Hs.486542.1.12	0.6919437	2.536469	0.22582555	0.089031465	0
ZAK	218833_at	Hs.444451.1.34	0.057304498	2.5166445	0.34323704	0.13638678	0
ZNF764	222120_at	Hs.132227.1.3	-0.29826826	2.5066853	0.3540722	0.14125116	0
CA5BP	238435_at	Hs.532326.1.10	1.2291596	2.5005007	0.15495674	0.061970286	0
MYNN	218926_at	Hs.507025.1.23	0.2847952	2.4942102	0.1963701	0.07873037	0
C21orf70	238703_at	Hs.410830.1.10	0.18405089	2.4900124	0.19011146	0.0763496	0
KIAA1430	226254_s_at	Hs.535734.1.4	0.43719223	2.480512	0.27963588	0.11273313	0
IGL@	215121_x_at	Hs.449585.1.72	0.19216146	2.474693	0.71076655	0.287214	0
MEF2A	208328_s_at	Hs.268675.1.28	-0.3278044	2.4712875	0.3359269	0.13593194	0
LASS6	242019_at	Hs.506829.1.15	0.05125723	2.4239419	0.17934169	0.07398762	0
RB1	211540_s_at	Hs.408528.1.2	-0.50281245	2.422527	0.19610132	0.080949076	0
TDGF1	206286_s_at	Hs.385870.1.17	0.23129295	2.418012	0.17669517	0.073074564	0
RECQL	205091_x_at	Hs.235069.1.4	-0.6272843	2.4115064	0.18591726	0.0770959	0
ATP6V1C1	202872_at	Hs.86905.1.21	0.98955184	2.4105282	0.24485399	0.1015769	0
TBC1D15	218268_at	Hs.284630.1.31	-0.5676096	2.4094605	0.22975665	0.095356055	0
CEP76	52285_f_at	Hs.236940.1.5	-0.14101326	2.4067073	0.3180874	0.13216704	0
DCP2	235258_at	Hs.443875.1.32	0.50562656	2.398262	0.19479293	0.081222534	0
TMEM50B	222907_x_at	Hs.433668.1.15	0.17013794	2.3941536	0.25262323	0.10551672	0
MOBP	207659_s_at	Hs.121333.1.11	0.22802109	2.387235	0.18249619	0.07644668	0
MSRB3	238583_at	Hs.339024.1.7	-0.57477266	2.3796642	0.24425459	0.10264246	0
TMEM107	235490_at	Hs.513933.1.4	-0.23550935	2.371401	0.3722347	0.15696827	0
HACE1	227471_at	Hs.434340.1.2	0.65855193	2.3679667	0.22445801	0.09478935	0
FBXO8	223240_at	Hs.76917.1.3	0.42706838	2.3657765	0.30179328	0.12756626	0
CLDN22	222738_at	Hs.333179.1.38	0.43539017	2.3647337	0.20864668	0.08823264	0
CTBS	218923_at	Hs.513557.1.5	-1.327741	2.3627613	0.2837435	0.12008979	0
CCDC112	235208_at	Hs.436121.1.4	0.5116455	2.3605099	0.3365373	0.14256975	0
TIFA	238858_at	Hs.310640.1.1	0.38839227	2.3592553	0.28607005	0.121254385	0

Table B.1 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
TROVE2	212839_s_at	Hs.288178.1.26	-1.0571882	2.3532507	0.36617494	0.15560387	0
LGR4	218326_s_at	Hs.502176.1.5	-0.76869583	2.329778	0.15107977	0.06484728	0.283459
RAB18	224377_s_at	Hs.406799.1.9	-0.9416127	2.3173647	0.3223788	0.13911441	0.280759
PML	235508_at	Hs.526464.1.16	-0.34548566	2.3065712	0.16857588	0.073085055	0.556222
ITGA11	222899_at	Hs.436416.1.2	-0.34872258	2.291071	0.1888094	0.08241098	0.551023
CSE1L	201111_at	Hs.90073.1.50	0.14871602	2.279355	0.28013745	0.122902066	0.545921
RPL13	229590_at	Hs.410817.1.15	-0.24679862	2.267775	0.18508166	0.081613764	0.535995
SNHG10	238691_at	Hs.448753.1.2	-0.39890036	2.2654715	0.29220653	0.12898266	0.535995
ZNF554	242864_at	Hs.307043.1.14	-0.104608364	2.2552607	0.18740857	0.083098404	0.526424
FDX1	203646_at	Hs.744.1.11	-0.67329955	2.2530272	0.442389	0.19635317	0.526424
ZNF839	219086_at	Hs.106005.1.22	-0.3923847	2.2382998	0.18055159	0.08066461	0.521765
NCAPH2	40640_at	Hs.180903.1.42	0.21577105	2.228244	0.32950342	0.14787582	0.517188
SSR1	200890_s_at	Hs.114033.1.18	0.5932451	2.224057	0.2962322	0.13319452	0.508271
LMAN2	200805_at	Hs.75864.1.4	0.5786093	2.221357	0.18985364	0.08546741	0.508271
GAS7	211067_s_at	Hs.462214.1.9	-0.22900966	2.202179	0.19470963	0.08841681	0.471676
YTHDC2	205836_s_at	Hs.231942.1.36	0.5074808	2.1897674	0.21334335	0.097427405	0.471676
MAPK1	208351_s_at	Hs.431850.1.20	0.1906501	2.1891923	0.18477798	0.08440463	0.471676
C16orf72	225183_at	Hs.221497.1.21	-0.31488162	2.1875696	0.13839531	0.063264415	0.471676
LRRC23	206076_at	Hs.155586.1.10	-0.63472086	2.183083	0.12569305	0.057575937	0.471676
LRRC23	206076_at	Hs.155586.1.11	-0.63282114	2.183083	0.12569305	0.057575937	0.471676
UBE2D3	200669_s_at	Hs.518773.1.8	0.38481426	2.1823866	0.30243814	0.13858138	0.471676
TCF3	210776_x_at	Hs.371282.1.10	-0.106136955	2.1784737	0.4517228	0.20735747	0.471676
CDC6	203967_at	Hs.405958.1.26	-0.2132555	2.177365	0.15716366	0.072180666	0.471676
GNL1	203307_at	Hs.83147.1.3	0.61722857	2.164654	0.16902089	0.078082174	0.460621
MCM10	222962_s_at	Hs.198363.1.34	-0.96631706	2.1644695	0.23010805	0.10631152	0.460621
TMEM37	1554485_s_at	Hs.26216.1.6	0.03977251	2.1634853	0.16524525	0.07637919	0.460621
PCYT1A	204210_s_at	Hs.435767.1.19	0.29847392	2.1467404	0.1556592	0.07250956	0.446663
PCYT1A	204210_s_at	Hs.435767.1.20	0.30018377	2.1467404	0.1556592	0.07250956	0.446663
CCDC90B	222577_at	Hs.368866.1.1	-0.7028938	2.1437013	0.16048053	0.07486142	0.446663
CCNB1	228729_at	Hs.23960.1.20	0.47262424	2.1429305	0.21005115	0.09802052	0.446663
PPP2R5C	201877_s_at	Hs.368264.1.51	-0.3943284	2.120882	0.16883385	0.07960549	0.436737
FAM82B	218549_s_at	Hs.145386.1.7	0.9419403	2.1182036	0.21005344	0.09916584	0.436737
ECHDC1	219974_x_at	Hs.486410.1.7	0.6870205	2.1176515	0.21449237	0.10128785	0.436737
USP12	215886_x_at	Hs.42400.1.21	-0.51895106	2.1138873	0.20513734	0.0970427	0.867052
UNKNOWN	216304_x_at	Hs.499145.1.10	-0.9454668	2.1079311	0.47460294	0.22515106	0.860723
NAP1L1	204528_s_at	Hs.524599.1.9	-0.5658632	2.0931306	0.19669592	0.093972124	0.842279
ZNF625	244406_at	Hs.512823.1.2	-0.08643651	2.0909638	0.27423894	0.13115433	0.842279
ZNF625	244406_at	Hs.512823.1.1	-0.08792645	2.0909638	0.27423894	0.13115433	0.842279
CREBBP	202160_at	Hs.459759.1.9	-0.32105392	2.087175	0.29624927	0.14193793	0.836305
CCDC112	235208_at	Hs.436121.1.2	0.509529	2.0800045	0.31663948	0.15223019	0.813235
CCNB1	228729_at	Hs.23960.1.22	0.47455522	2.073838	0.21953249	0.10585807	0.813235
CLDN8	214598_at	Hs.162209.1.2	0.166798	2.072558	0.12855011	0.062024858	0.813235
RAB3C	242328_at	Hs.482173.1.11	0.46679878	2.0719929	0.19453031	0.09388561	0.813235
LGR4	218326_s_at	Hs.502176.1.4	-0.77103645	2.0670462	0.12825878	0.062049307	0.807664
THUMPD3	223804_s_at	Hs.443081.1.23	0.21892972	2.044571	0.15693083	0.0767549	0.770712
MED14	202610_s_at	Hs.407604.1.7	1.2856821	2.0444512	0.1921122	0.09396762	0.770712
CALCOCO2	235076_at	Hs.514920.1.38	-0.19045748	2.042679	0.21938342	0.10739984	0.770712
SUGT1	223330_s_at	Hs.281902.1.25	-0.4994295	2.0421216	0.16912481	0.08281819	0.770712
NIPBL	212483_at	Hs.481927.1.71	0.45589712	2.0375817	0.19741976	0.09688925	0.770712
TFPI	209676_at	Hs.516578.1.9	0.071348496	2.0370169	0.29972816	0.14714073	0.770712
TFPI	209676_at	Hs.516578.1.6	0.06993606	2.0370169	0.29972816	0.14714073	0.770712
SLC35B3	222691_at	Hs.285847.1.7	0.599562	2.029042	0.32038867	0.15790145	0.765708
SGCB	226112_at	Hs.438953.1.3	0.34529564	2.0245678	0.13577081	0.067061625	0.760768
C10orf119	217905_at	Hs.124246.1.4	-0.83859515	2.0190372	0.27325416	0.13533884	0.755891
RERE	221643_s_at	Hs.463041.1.5	-2.0146828	2.0123615	0.1395951	0.0693688	1.126615
XRN1	1555785_a_at	Hs.435103.1.4	0.27524534	2.0037153	0.15646401	0.07808695	1.119484
FAM82B	218549_s_at	Hs.145386.1.13	0.9458413	1.9959928	0.20527661	0.102844365	1.091843
SMC2	213253_at	Hs.119023.1.48	1.132616	1.9918512	0.15927112	0.07996135	1.091843
NDFIP2	224802_at	Hs.525093.1.22	-0.48808166	1.9909753	0.15734121	0.079027206	1.091843
C10orf72	235471_at	Hs.522928.1.3	-0.9092554	1.98896	0.1581943	0.07953619	1.091843
LAS1L	235541_at	Hs.522675.1.13	1.3548659	1.9801536	0.17472735	0.08823929	1.085144
ZNF583	1553221_at	Hs.146854.1.13	-0.035602402	1.964537	0.17502218	0.0890908	1.078528
ZNF440	241731_x_at	Hs.418192.1.5	-0.09103841	1.9575459	0.1562666	0.07982781	1.071991
ACTR3	213102_at	Hs.433512.1.22	0.036977317	1.9517193	0.27273524	0.13974102	1.065533
PLSCR1	202430_s_at	Hs.130759.1.13	0.27680248	1.9449422	0.21825728	0.11221787	1.059153

Table B.2 Significantly Shortened Genes in RARS Subtype of MDS

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
SLC48A1	48106_at	Hs.438867.1.13	-0.61799586	5.7372093	0.5755558	0.10031982	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.8216926	5.648613	0.90877783	0.16088513	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.82445365	5.648613	0.90877783	0.16088513	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.8275878	5.648613	0.90877783	0.16088513	0
GTPBP8	223486_at	Hs.127496.1.12	0.2494566	5.561915	0.43004048	0.07731878	0
YARS2	218470_at	Hs.505231.1.6	-0.62237966	5.4671216	0.7740967	0.14159128	0
SF4	209547_s_at	Hs.515274.1.3	-0.08024067	5.4530296	0.60892045	0.11166645	0
YARS2	218470_at	Hs.505231.1.4	-0.62435335	5.42969	0.72407186	0.13335419	0
C4A	214428_x_at	Hs.534847.1.59	0.61680543	5.1720815	0.4459478	0.08622211	0
ATP7A	205198_s_at	Hs.496414.1.21	1.4367559	5.104328	0.52394885	0.10264795	0
ADH5	208847_s_at	Hs.78989.1.10	0.380158	5.092344	0.48532033	0.09530392	0
C9orf85	238579_at	Hs.534190.1.18	1.0761693	5.0084906	0.52041316	0.103906184	0
RTN1	203485_at	Hs.368626.1.2	-0.43394294	4.750045	0.41961932	0.088340074	0
FLII	212024_x_at	Hs.513984.1.1	-0.22954866	4.7227907	0.626297	0.13261163	0
TGM2	211573_x_at	Hs.517033.1.19	0.14173996	4.6730127	0.25989664	0.055616505	0
POLR2C	216282_x_at	Hs.79402.1.16	-0.28841537	4.6718	0.6423373	0.13749246	0
GUF1	218884_s_at	Hs.546419.1.31	0.33731115	4.613265	0.3584781	0.07770594	0
MYL4	217274_x_at	Hs.463300.1.13	-0.20247401	4.5670924	0.41655365	0.09120762	0
SNX11	53912_at	Hs.15827.1.19	-0.19781709	4.5416946	0.61924314	0.13634628	0
LYPLA2	215568_x_at	Hs.533479.1.18	-1.7725393	4.4841948	0.54319966	0.1211365	0
TCF3	210776_x_at	Hs.371282.1.10	-0.1084726	4.4735336	0.7774922	0.17379823	0
SCARB2	201647_s_at	Hs.349656.1.10	0.36133552	4.4708824	0.32764408	0.073283985	0
C9orf82	231995_at	Hs.178357.1.3	1.0583493	4.429821	0.34199116	0.07720203	0
INCENP	219769_at	Hs.142179.1.21	-0.7378959	4.393324	0.4531498	0.10314509	0
IGL@	215121_x_at	Hs.449585.1.72	0.189016	4.3703465	1.4063442	0.32179236	0
APIP	218698_at	Hs.447794.1.3	-0.7502069	4.3698025	0.58501315	0.13387634	0
MYL4	216054_x_at	Hs.463300.1.13	-0.20097764	4.3590717	0.51268554	0.117613465	0
RAB3C	242328_at	Hs.482173.1.9	0.46809	4.3265853	0.4088005	0.094485715	0
GAN	220124_at	Hs.112569.1.10	-0.257992	4.3021965	0.27824825	0.06467586	0
NCAPH2	40640_at	Hs.180903.1.42	0.21494682	4.2921157	0.56252706	0.13106056	0
TPI1	210050_at	Hs.524219.1.15	-0.6472954	4.2864547	0.6089015	0.14205247	0
TMEM50B	222907_x_at	Hs.433668.1.15	0.16744584	4.232169	0.39638108	0.09365908	0
ACTL6A	202666_s_at	Hs.435326.1.30	0.28875592	4.22381	0.42004395	0.09944669	0
NIP30	224248_x_at	Hs.396740.1.5	-0.29330036	4.192217	0.46940923	0.1119716	0
SCARB2	201647_s_at	Hs.349656.1.9	0.35963616	4.125346	0.270384	0.06554214	0
MAP1D	1569029_at	Hs.298250.1.17	0.054778785	4.124055	0.3425103	0.08305183	0
VTI1A	235034_at	Hs.194554.1.28	-0.85597545	4.102327	0.27850726	0.06789007	0
ENOX2	234003_at	Hs.171458.1.30	1.8543383	4.1019716	0.32428247	0.079055265	0
LARS	222428_s_at	Hs.432674.1.6	0.5440337	4.0562267	0.49676812	0.1224705	0
C22orf25	235396_at	Hs.474233.1.27	0.18426369	4.015626	0.37731802	0.093962446	0
ZAK	218833_at	Hs.444451.1.34	0.05753051	4.0027933	0.51752543	0.12929107	0
C10orf119	217905_at	Hs.124246.1.4	-0.84148204	3.9752553	0.4727038	0.118911564	0
CFLAR	210564_x_at	Hs.390736.1.25	0.08189853	3.896275	0.41506824	0.106529504	0
CFLAR	210564_x_at	Hs.390736.1.24	0.080246255	3.896275	0.41506824	0.106529504	0
CCDC112	235208_at	Hs.436121.1.2	0.50958604	3.880592	0.5236501	0.13494077	0
CYP20A1	219565_at	Hs.446065.1.25	0.089794144	3.868577	0.35865086	0.09270873	0
CYP20A1	219565_at	Hs.446065.1.24	0.088258564	3.868577	0.35865086	0.09270873	0
TMBIM4	219206_x_at	Hs.505934.1.8	-0.58123636	3.774348	0.5062448	0.13412774	0
BCLAF1	214499_s_at	Hs.486542.1.16	0.69588757	3.774111	0.47461712	0.12575601	0
RB1	211540_s_at	Hs.408528.1.2	-0.5059015	3.7365422	0.24559449	0.06572774	0
BCLAF1	201083_s_at	Hs.486542.1.12	0.69331974	3.7244046	0.2387613	0.06410724	0
SET	213047_x_at	Hs.436687.1.19	1.2046957	3.7098792	0.29839754	0.08043323	0
CCDC112	235208_at	Hs.436121.1.4	0.5114242	3.6900892	0.49350268	0.13373733	0
HRH1	205579_at	Hs.1570.1.7	0.2213243	3.6776373	0.21311162	0.05794797	0
TMEM37	1554485_s_at	Hs.26216.1.6	0.04074107	3.6624076	0.24226198	0.06614828	0
ZKSCAN5	203731_s_at	Hs.110839.1.21	0.78611934	3.6466186	0.23775017	0.06519743	0
POLR2C	216282_x_at	Hs.79402.1.17	-0.2872314	3.6224988	0.51468384	0.14207979	0
PHKB	202739_s_at	Hs.78060.1.71	-0.2991096	3.6115365	0.26298296	0.072817475	0
RAB3C	242328_at	Hs.482173.1.11	0.4697978	3.5982	0.306933	0.08530181	0
LDB1	35160_at	Hs.454418.1.3	-0.8586379	3.580287	0.40390825	0.11281449	0
LOC157627	214839_at	Hs.12513.1.2	0.8514964	3.5749319	0.250135	0.06996917	0
CEP76	52285_f_at	Hs.236940.1.5	-0.14467315	3.5537412	0.42216283	0.11879391	0
CHEK1	205394_at	Hs.24529.1.27	-0.6537733	3.5509074	0.26477623	0.0745658	0
POLR2C	216282_x_at	Hs.79402.1.18	-0.28570497	3.528541	0.49086607	0.13911304	0
NSDHL	215093_at	Hs.57698.1.21	2.1401346	3.5255098	0.43529338	0.12346963	0
MRP63	204387_x_at	Hs.458367.1.3	-0.53065604	3.5034926	0.43094385	0.12300407	0
UNKNOWN	2163B04_x_at	Hs.499145.1.10	-0.9465659	3.499832	0.64193416	0.18341857	0

Table B.2 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
C21orf70	238703_at	Hs.410830.1.10	0.18146467	3.4950607	0.22135493	0.063333645	0
LGI4	242670_at	Hs.65256.1.20	-0.07450341	3.4210727	0.195436	0.05712711	0
ECHDC1	219974_x_at	Hs.486410.1.6	0.68597746	3.376469	0.29584098	0.087618455	0
RPL13	229590_at	Hs.410817.1.15	-0.24883758	3.3725607	0.21269386	0.06306598	0
DNPEP	38703_at	Hs.258551.1.5	0.09790791	3.359765	0.2472502	0.073591515	0
CFLAR	210564_x_at	Hs.390736.1.14	0.078648955	3.3200064	0.35124382	0.10579613	0
MBNL3	219814_at	Hs.105134.1.8	1.907791	3.3199637	0.361925	0.10901475	0
HMOX2	218121_at	Hs.284279.1.14	-0.31814796	3.3113642	0.32487988	0.09811059	0
NNT	202783_at	Hs.482043.1.62	0.46198818	3.3102596	0.4045999	0.12222604	0
SDF4	217855_x_at	Hs.42806.1.2	-2.3330908	3.2848952	0.33102554	0.10077202	0
ITPR3	201188_s_at	Hs.65758.1.102	0.6226233	3.2840903	0.62625074	0.1906923	0
MEF2A	208328_s_at	Hs.268675.1.28	-0.32830226	3.279225	0.3742845	0.1141381	0
LCP2	244251_at	Hs.304475.1.1	0.56058353	3.277321	0.20837808	0.06358183	0
SNHG10	238691_at	Hs.448753.1.2	-0.40035558	3.2751172	0.38737178	0.11827723	0
SSR1	200890_s_at	Hs.114033.1.18	0.59103644	3.2630973	0.3584236	0.109841526	0
GJA4	40687_at	Hs.296310.1.12	-1.5591948	3.2557175	0.28271466	0.08683636	0
TBCEL	231997_at	Hs.504136.1.20	-0.65814775	3.2375016	0.17639105	0.054483693	0
KCNN1	206231_at	Hs.158173.1.9	-0.08177085	3.2067187	0.24171028	0.075376205	0
ZNF583	1553221_at	Hs.146854.1.13	-0.03825071	3.1976757	0.2363683	0.07391878	0
LAS1L	235541_at	Hs.522675.1.13	1.3761423	3.173638	0.21037844	0.06628936	0
VPS4B	218171_at	Hs.126550.1.2	-0.12559691	3.171875	0.31896913	0.10056169	0
NEFL	221801_x_at	Hs.521461.1.2	0.8733962	3.1633542	0.3114264	0.098448165	0
TMEM33	218465_at	Hs.31082.1.28	0.3343565	3.1625812	0.20386416	0.06446132	0
SSX2IP	203019_x_at	Hs.22587.1.2	-1.3182567	3.1514456	0.25975174	0.082423046	0
CDC6	203967_at	Hs.405958.1.26	-0.2183671	3.1490893	0.20087172	0.063787244	0
MCM10	222962_s_at	Hs.198363.1.34	-0.97089744	3.1413202	0.28662568	0.0912437	0
LASS6	242019_at	Hs.506829.1.15	0.05188184	3.1350112	0.20840515	0.06647669	0
ORC2L	204853_at	Hs.444870.1.4	0.07740553	3.1313543	0.16965747	0.05418022	0
LGR4	218326_s_at	Hs.502176.1.5	-0.77345484	3.1284075	0.17393205	0.055597633	0
TCF3	213730_x_at	Hs.371282.1.8	-0.111532636	3.097432	0.4522369	0.14600383	0
C14orf93	219009_at	Hs.255874.1.4	-0.46901742	3.0947025	0.21513459	0.06951705	0
TAF3	235119_at	Hs.527688.1.15	-0.9805883	3.0694587	0.3459357	0.11270251	0
CLU	222043_at	Hs.436657.1.7	0.88005	3.060298	0.41567755	0.1358291	0
LLPH	224446_at	Hs.504820.1.6	-0.58323675	3.0488029	0.33825606	0.11094718	0
KLF3	222913_at	Hs.298658.1.19	0.32831696	3.0384748	0.29153335	0.095947266	0
EIF5B	201025_at	Hs.158688.1.67	0.02708885	3.034767	0.40338314	0.13292064	0
PRPS1	208447_s_at	Hs.56.1.15	1.5183418	3.033368	0.18315496	0.060380064	0
SNX3	200067_x_at	Hs.12102.1.3	0.6697267	3.0042999	0.23455179	0.07807203	0
TCF3	213730_x_at	Hs.371282.1.10	-0.10989659	2.973872	0.41908586	0.14092264	0
METTL6	1553689_s_at	Hs.149487.1.8	0.22278166	2.9719586	0.14032987	0.047217976	0
KIAA1704	223606_x_at	Hs.507922.1.17	-0.51758164	2.97162	0.19359457	0.065147825	0
KIAA1704	223606_x_at	Hs.507922.1.22	-0.5157831	2.97162	0.19359457	0.065147825	0
CSTF1	32723_at	Hs.172865.1.25	0.15818629	2.9654145	0.33624685	0.11338949	0
TROVE2	212839_s_at	Hs.288178.1.26	-1.0650917	2.962275	0.37901962	0.12794882	0
MEF2A	208328_s_at	Hs.268675.1.29	-0.32689106	2.9568112	0.3255732	0.11010957	0
CA5BP	238435_at	Hs.532326.1.10	1.244812	2.9538028	0.15540123	0.05261056	0
SLC46A3	214719_at	Hs.117167.1.1	-0.5248119	2.9518797	0.3382423	0.11458539	0
IQCK	215131_at	Hs.460217.1.5	-0.31369156	2.943969	0.5908432	0.20069614	0
NAP1L1	204528_s_at	Hs.524599.1.9	-0.57554585	2.9205835	0.2527877	0.086553834	0
ZNF625	244406_at	Hs.512823.1.2	-0.089313634	2.9170747	0.335042	0.114855476	0
ZNF625	244406_at	Hs.512823.1.1	-0.090789475	2.9170747	0.335042	0.114855476	0
MED14	202610_s_at	Hs.407604.1.7	1.3020349	2.9132266	0.25071305	0.08606026	0
ZNF280D	1568951_at	Hs.511477.1.25	-0.3678428	2.9100296	0.19188254	0.065938346	0
RPP30	1556061_at	Hs.139120.1.36	-0.8700775	2.90301	0.15271026	0.052604113	0
CUL4B	215997_s_at	Hs.102914.1.20	1.7343308	2.893258	0.13355622	0.046161182	0
MED28	224416_s_at	Hs.434075.1.9	0.31812426	2.88288	0.35875237	0.124442354	0
DKC1	216212_s_at	Hs.4747.1.24	2.3544176	2.8784256	0.17265996	0.059984166	0
FAM82B	218549_s_at	Hs.145386.1.7	0.938415	2.8725727	0.22277784	0.07755342	0
SAG	206671_at	Hs.32721.1.21	0.102415755	2.8715951	0.21892166	0.07623695	0
NDFIP2	224802_at	Hs.525093.1.22	-0.49185902	2.8690903	0.16485944	0.057460528	0
NOP56	200875_s_at	Hs.376064.1.37	0.11114907	2.863022	0.15891087	0.05550459	0
C11orf24	52164_at	Hs.303025.1.2	-0.7301278	2.850937	0.20833191	0.07307489	0
UBXN6	220757_s_at	Hs.435255.1.5	-0.10443067	2.8243673	0.18002169	0.06373876	0
UBE2D3	200669_s_at	Hs.518773.1.8	0.3834408	2.8033655	0.32690334	0.11661103	0
PCYT1A	204210_s_at	Hs.435767.1.20	0.3001241	2.782404	0.17168432	0.061703596	0
PCYT1A	204210_s_at	Hs.435767.1.19	0.29828176	2.782404	0.17168432	0.061703596	0
MAF	1566324_a_at	Hs.134859.1.5	-0.26141527	2.7614222	0.19379465	0.07017929	0

Table B.2 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
NEFH	204412_s_at	Hs.198760.1.10	0.19851239	2.7476187	0.25263622	0.09194734	0
ECHDC1	219974_x_at	Hs.486410.1.7	0.6884233	2.7408304	0.22193559	0.08097385	0
TRIB2	202479_s_at	Hs.467751.1.23	-0.028318256	2.7302709	0.23944554	0.087700285	0
ITGA11	222899_at	Hs.436416.1.2	-0.3468688	2.7257805	0.19240928	0.07058869	0
KIAA1430	226254_s_at	Hs.535734.1.4	0.4351487	2.7125144	0.23310706	0.085937634	0
LMAN2	200805_at	Hs.75864.1.4	0.5749902	2.7014	0.21042576	0.077895075	0
C12orf4	218374_s_at	Hs.302977.1.4	-0.6495363	2.686116	0.17260307	0.06425749	0
TFPI2	209277_at	Hs.438231.1.11	0.7697263	2.677815	0.16121364	0.060203426	0
SNX3	213545_x_at	Hs.12102.1.3	0.67212325	2.6704524	0.2776519	0.10397186	0
CETN1	207209_at	Hs.122511.1.1	-0.15516616	2.663901	0.15836585	0.059448846	0
CD5L	206680_at	Hs.134035.1.3	-1.1762269	2.6562011	0.23734629	0.08935554	0
THUMPD3	223804_s_at	Hs.443081.1.23	0.21813832	2.652941	0.17289087	0.06516951	0
SNX13	1553148_a_at	Hs.487648.1.1	0.73132414	2.6201115	0.34259748	0.13075684	0
NPL	221210_s_at	Hs.496969.1.23	-1.0769951	2.5994887	0.2571982	0.09894185	0
MED4	222438_at	Hs.181112.1.5	-0.5078529	2.5962334	0.36934936	0.14226355	0
AK2	205996_s_at	Hs.470907.1.3	-1.592798	2.5944767	0.3229493	0.124475695	0
GNL1	203307_at	Hs.83147.1.3	0.61480165	2.5943959	0.17294818	0.066662215	0
MYNN	218926_at	Hs.507025.1.23	0.28409365	2.5897236	0.178716	0.069009684	0
EBNA1BP2	201323_at	Hs.346868.1.2	-1.465065	2.5891237	0.30324656	0.11712324	0
FGB	204988_at	Hs.300774.1.15	0.41340473	2.5774174	0.16352487	0.06344524	0
WDTC1	215497_s_at	Hs.469154.1.38	-1.6707537	2.576859	0.16669345	0.064688616	0
C1orf107	214193_s_at	Hs.194754.1.29	-1.0307766	2.5670187	0.20402503	0.07947937	0
TBC1D15	218268_at	Hs.284630.1.31	-0.5775118	2.5508893	0.19494772	0.07642344	0
CSTF1	32723_at	Hs.172865.1.23	0.15676913	2.542817	0.2634377	0.10360072	0
RGS10	204316_at	Hs.501200.1.2	-0.8443574	2.5397706	0.39002454	0.15356684	0
CREM	214508_x_at	Hs.200250.1.29	-0.93396354	2.537333	0.1841152	0.07256249	0
BMP7	211259_s_at	Hs.473163.1.16	0.15980645	2.5365777	0.16025454	0.06317746	0
BMP7	211259_s_at	Hs.473163.1.17	0.16120863	2.5365777	0.16025454	0.06317746	0
PCMT1	208857_s_at	Hs.279257.1.17	0.70063204	2.5337167	0.21031654	0.08300713	0
FAM82B	218549_s_at	Hs.145386.1.13	0.9418029	2.5241895	0.20496678	0.08120103	0
UNKNOWN	235386_at	Hs.400256.1.2	0.31199726	2.520793	0.1635524	0.06488133	0
UNKNOWN	235386_at	Hs.400256.1.1	0.31042647	2.520793	0.1635524	0.06488133	0
ECHDC1	219974_x_at	Hs.486410.1.5	0.6832774	2.5128727	0.21839458	0.08691032	0
ATP5S	1554177_a_at	Hs.438489.1.12	-0.4548137	2.5098436	0.36407876	0.14506035	0
ZNF764	222120_at	Hs.132227.1.3	-0.30075693	2.5090635	0.29001474	0.11558685	0
TRAPPC10	209412_at	Hs.126221.1.44	0.17810908	2.5088394	0.23464853	0.09352872	0
TDGF1	206286_s_at	Hs.385870.1.17	0.23056006	2.5060644	0.15812871	0.06309842	0
CREBBP	202160_at	Hs.459759.1.9	-0.3225065	2.4829473	0.29017234	0.11686609	0
CCNB1	228729_at	Hs.23960.1.22	0.47811082	2.4767897	0.21293855	0.085973606	0
SLC35B3	222691_at	Hs.285847.1.7	0.59748983	2.4749522	0.2556907	0.10331137	0
ESF1	218859_s_at	Hs.369284.1.19	0.123597056	2.4717917	0.2477979	0.10025032	0
PTPRA	213795_s_at	Hs.269577.1.58	0.11266252	2.464157	0.17493701	0.07099263	0
PTPRA	213795_s_at	Hs.269577.1.59	0.114365935	2.464157	0.17493701	0.07099263	0
RHO	206454_s_at	Hs.247565.1.15	0.26344526	2.463409	0.15914886	0.06460513	0
TRAPPC10	209412_at	Hs.126221.1.45	0.17978275	2.431319	0.23035544	0.09474505	0
LOC440149	236366_at	Hs.41423.1.1	-0.4826244	2.4286003	0.11920655	0.049084466	0
ADRBK1	38447_at	Hs.83636.1.51	-0.73271865	2.4247313	0.24397027	0.100617446	0
CTBS	218923_at	Hs.513557.1.5	-1.3329916	2.4239357	0.24242449	0.10001276	0
LRRC23	206076_at	Hs.155586.1.10	-0.6450322	2.4025657	0.13032213	0.054242898	0
LRRC23	206076_at	Hs.155586.1.11	-0.64279926	2.4025657	0.13032213	0.054242898	0
CDC6	203968_s_at	Hs.405958.1.22	-0.2169064	2.3952513	0.19093764	0.07971507	0
SRP72	208802_at	Hs.237825.1.39	0.34916636	2.38225	0.15546304	0.065258905	0
40067	201308_s_at	Hs.128199.1.37	0.36493343	2.3798862	0.25781578	0.10833114	0
METTL1	204027_s_at	Hs.42957.1.2	-0.5970647	2.3792074	0.17637381	0.07413133	0
MOBP	207659_s_at	Hs.121333.1.11	0.2273827	2.3782814	0.15946972	0.067052506	0
C18orf25	217508_s_at	Hs.116486.1.13	-0.13175452	2.3779442	0.19079655	0.08023593	0
ZNF554	242864_at	Hs.307043.1.14	-0.10706727	2.3712137	0.21351027	0.09004261	0
PGBD1	235411_at	Hs.144527.1.13	0.6123347	2.3705547	0.19683057	0.08303144	0
ATP6V1C1	202872_at	Hs.86905.1.21	0.9887398	2.3662436	0.18341316	0.077512376	0
VASN	225867_at	Hs.372579.1.7	-0.3195288	2.364686	0.1322135	0.055911653	0
LRRCC1	231872_at	Hs.193115.1.27	0.9352016	2.3643148	0.24170983	0.10223251	0
TET2	235461_at	Hs.367639.1.10	0.3850952	2.3637135	0.35688752	0.15098594	0
RIOK1	224450_s_at	Hs.437474.1.29	0.59296566	2.3619792	0.18361205	0.07773652	0
TPRG1	229764_at	Hs.338851.1.15	0.29519525	2.3553226	0.14890897	0.06322233	0
TFPI	209676_at	Hs.516578.1.9	0.071784325	2.3546596	0.28234172	0.11990766	0
TFPI	209676_at	Hs.516578.1.6	0.07042374	2.3546596	0.28234172	0.11990766	0

Table B.2 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
WIPI2	214699_x_at	Hs.122363.1.30	0.7212605	2.352619	0.18650159	0.079274036	0
RABGEF1	218310_at	Hs.530053.1.33	0.7560498	2.3491335	0.30550563	0.13005035	0
KLHL24	242088_at	Hs.407709.1.19	0.29175964	2.3477008	0.23989213	0.102181725	0
SLC16A1	209900_s_at	Hs.75231.1.3	-1.2595576	2.346784	0.25778508	0.1098461	0
SEC22A	218703_at	Hs.477361.1.19	0.25403088	2.3323953	0.21045613	0.09023176	0
LETM1	218939_at	Hs.120165.1.4	0.30406404	2.3295457	0.12243155	0.052555975	0
C19orf42	219097_x_at	Hs.356467.1.24	-0.08347755	2.31215	0.19568455	0.08463316	0
C16orf72	225183_at	Hs.221497.1.21	-0.31668898	2.3004384	0.13328874	0.057940584	0.237627
TMEM50B	222907_x_at	Hs.433668.1.14	0.16560616	2.2978344	0.20603943	0.08966679	0.237627
RAB12	235059_at	Hs.270074.1.10	-0.14796111	2.2935717	0.21029675	0.09168963	0.237627
MAPK13	210059_s_at	Hs.178695.1.25	0.6270971	2.2934241	0.18736285	0.08169568	0.237627
FDX1	203646_at	Hs.744.1.11	-0.68035144	2.2933164	0.38471687	0.1677557	0.237627
MAP2K5	211370_s_at	Hs.114198.1.23	-0.34860492	2.2920635	0.17145854	0.07480532	0.237627
DSTN	201021_s_at	Hs.304192.1.3	0.12803051	2.2912254	0.15423757	0.06731663	0.237627
SPP1	209875_s_at	Hs.313.1.13	0.37346962	2.291148	0.13998412	0.06109781	0.237627
PTBP1	212016_s_at	Hs.172550.1.28	-0.11301867	2.2833903	0.110503174	0.048394345	0.234386
POSTN	210809_s_at	Hs.136348.1.11	-0.51942664	2.2820232	0.31103408	0.13629751	0.234386
APP	200602_at	Hs.434980.1.2	0.1626046	2.2792597	0.16988528	0.07453529	0.234386
YTHDC2	205836_s_at	Hs.231942.1.36	0.5079158	2.2708077	0.18650076	0.0821297	0.231233
CREB1	204314_s_at	Hs.516646.1.28	0.09304143	2.2691777	0.19669974	0.086683266	0.231233
LOC440131	233804_at	Hs.132371.1.1	-0.52278405	2.2675588	0.09992288	0.04406628	0.231233
PML	235508_at	Hs.526464.1.16	-0.34372014	2.2476883	0.1429811	0.06361251	0.230201
PROC	206259_at	Hs.224698.1.14	0.042395297	2.2437317	0.11084641	0.049402703	0.227158
NTS	206291_at	Hs.80962.1.6	-0.5656396	2.2407854	0.09933486	0.044330377	0.227158
C10orf84	218390_s_at	Hs.372309.1.3	-0.8472347	2.240104	0.18093759	0.08077196	0.227158
FLJ39582	239015_at	Hs.517430.1.1	0.18581876	2.2117777	0.16795003	0.07593441	0.224195
AGBL5	231857_s_at	Hs.138207.1.20	-0.012814504	2.2095487	0.11542818	0.052240614	0.224195
CCNG2	202769_at	Hs.13291.1.25	0.36670566	2.209475	0.27806687	0.125852	0.224195
NIPBL	212483_at	Hs.481927.1.71	0.45627663	2.2059407	0.1757034	0.079650104	0.222263
SRRT	201679_at	Hs.111801.1.18	0.7941543	2.2031438	0.15067568	0.06839122	0.222263
CLDN22	222738_at	Hs.333179.1.38	0.43344855	2.193427	0.18617707	0.08487954	0.218496
UNKNOWN	231820_x_at	Hs.288995.1.15	-0.036654234	2.1913295	0.33292973	0.15193048	0.218496
CNBP	206158_s_at	Hs.518249.1.3	0.25873283	2.1893034	0.22228867	0.10153397	0.218496
ISY1	223429_x_at	Hs.512661.1.7	0.25709602	2.1876268	0.15288997	0.0698885	0.218496
MGAT2	203102_s_at	Hs.93338.1.10	-0.460216	2.179824	0.2214526	0.10159195	0.216659
SPOCK3	235342_at	Hs.481133.1.3	0.41712886	2.177764	0.1917732	0.088059686	0.216659
DENR	238982_at	Hs.22393.1.3	-0.5449772	2.173339	0.18518159	0.08520603	0.215753
AMOT	209521_s_at	Hs.528051.1.2	1.6303144	2.1687732	0.3432455	0.15826713	0.213078
RHOB	212099_at	Hs.502876.1.4	-0.023604326	2.1561992	0.17599672	0.08162359	0.212201
TMEM33	218465_at	Hs.31082.1.27	0.3330353	2.1502845	0.1218869	0.05668408	0.210469
JRKL	206734_at	Hs.105940.1.2	-0.69804835	2.147061	0.18758225	0.08736698	0.210469
CCNB1	228729_at	Hs.23960.1.20	0.4757396	2.141789	0.16881964	0.078821786	0.209614
GNB5	207124_s_at	Hs.155090.1.10	-0.3711019	2.1118398	0.115127236	0.05451514	0.414177
HSPA4	211016_x_at	Hs.90093.1.1	0.5237326	2.1107209	0.13732162	0.06505911	0.414177
IL20RA	222829_s_at	Hs.445868.1.3	0.69811755	2.1099763	0.09836775	0.04662031	0.414177
TSR1	239042_at	Hs.388170.1.1	-0.24620183	2.0929139	0.16346776	0.078105345	0.410877
FAM5B	214822_at	Hs.495918.1.7	-1.130242	2.091177	0.08952668	0.042811625	0.410877
HACE1	227471_at	Hs.434340.1.2	0.66061795	2.0811572	0.16517192	0.079365425	0.407629
ETAA1	219216_at	Hs.353022.1.14	0.008217451	2.0805902	0.25891614	0.1244436	0.407629
ACOX3	204241_at	Hs.479122.1.6	0.3147885	2.0740707	0.17783216	0.08574064	0.406024
LRRC41	1558983_at	Hs.144941.1.29	-1.4533498	2.059942	0.1618272	0.07855911	0.402852
KLHL20	210634_at	Hs.495035.1.12	-1.1402122	2.0563173	0.14375314	0.06990805	0.402852
RPS6KB1	204171_at	Hs.463642.1.34	-0.17887636	2.0439336	0.2055167	0.10054959	0.399729
RECQL	205091_x_at	Hs.235069.1.4	-0.63719964	2.043589	0.13234714	0.06476211	0.399729
FBXO43	236852_at	Hs.339577.1.2	0.9850915	2.0345485	0.17219889	0.084637396	0.594981
SARDH	211322_s_at	Hs.198003.1.19	1.2181202	2.0319674	0.15852779	0.0780169	0.594981
SCAMP1	206668_s_at	Hs.482587.1.20	0.4867733	2.0244153	0.13480511	0.06658965	0.590439
OSBPL6	238575_at	Hs.318775.1.24	0.06035313	2.0232143	0.14258036	0.070472196	0.590439
ABHD4	242023_at	Hs.445665.1.23	-0.47330984	2.015077	0.12395036	0.061511472	0.588194
CORO2A	205538_at	Hs.113094.1.5	1.1247295	2.0063038	0.11312505	0.056384806	0.572945
SRP72	208800_at	Hs.237825.1.32	0.3455995	2.004491	0.16442129	0.08202645	0.572945
CHST5	64900_at	Hs.156784.1.13	-0.26481235	2.0028749	0.2016998	0.10070514	0.572945
ZNF440	241731_x_at	Hs.418192.1.5	-0.093665265	2.0007522	0.11839804	0.059176765	0.572945
TIMM8A	205217_at	Hs.447877.1.4	1.4620123	1.9994783	0.12655878	0.0632959	0.572945
ACTR3	213102_at	Hs.433512.1.22	0.037726484	1.9993057	0.22085416	0.11046543	0.572945
C14orf132	218820_at	Hs.6434.1.11	-0.39855245	1.997425	0.09989846	0.050013624	0.572945
TNFAIP1	201208_s_at	Hs.76090.1.18	-0.22651432	1.9944954	0.10680619	0.05355048	0.568732

Table B.2 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
MAP3K7IP3	227357_at	Hs.188256.1.3	1.2861165	1.9905231	0.1971885	0.09906366	0.568732
RGL3	1556355_x_at	Hs.375142.1.3	-0.09531541	1.985967	0.24961174	0.12568775	0.566649
MFAP3	214588_s_at	Hs.432818.1.12	0.5525042	1.9759645	0.1676259	0.084832445	0.564581
UNKNOWN	237787_at	Hs.202533.1.4	-1.2397107	1.9639425	0.111597165	0.05682303	0.750037

Table B.3 Significantly Lengthened Genes in RARS Subtype of MDS

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
MTHFD2	201761_at	Hs.469030.1.17	0.01359339	-3.9015455	-0.2860843	0.07332589	0
PABPC3	208113_x_at	Hs.458280.1.1	-0.52876973	-3.8701887	-0.67753804	0.17506589	0
MRP63	204386_s_at	Hs.458367.1.2	-0.53289753	-3.5733118	-0.30665198	0.0858173	0.21307881
STMN1	200783_s_at	Hs.209983.1.5	-1.7169414	-3.5146925	-0.42143488	0.11990661	0.21307881

Table B.4 Significantly Altered Genes in RAEB1 Subtype of MDS

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
SLC46A3	214719_at	Hs.117167.1.1	-0.4888284	5.3524456	0.6942899	0.12971452	0
ZNF554	242864_at	Hs.307043.1.14	-0.081884824	4.231268	0.3777523	0.08927639	0
ZAK	218833_at	Hs.444451.1.34	0.07552161	4.176079	0.6293019	0.15069206	0
ZNF625	244406_at	Hs.512823.1.1	-0.06645208	4.1583734	0.5328181	0.12813137	0
ZNF625	244406_at	Hs.512823.1.2	-0.06485978	4.1583734	0.5328181	0.12813137	0
SPOCK3	235342_at	Hs.481133.1.3	0.43105412	3.937637	0.44990376	0.11425729	0
C9orf85	238579_at	Hs.534190.1.18	1.0749497	3.8444207	0.39695084	0.103253745	0
SCARB2	201647_s_at	Hs.349656.1.10	0.3737547	3.7910337	0.32421562	0.08552169	0
ZKSCAN5	203731_s_at	Hs.110839.1.21	0.7864074	3.760412	0.2774588	0.07378414	0
C9orf82	231995_at	Hs.178357.1.3	1.056682	3.7570972	0.35759392	0.09517825	0
LYPLA2	215568_x_at	Hs.533479.1.18	-1.7664552	3.747811	0.5598092	0.14936964	0
SCARB2	201647_s_at	Hs.349656.1.9	0.37218907	3.6987777	0.28874964	0.07806623	0
FLII	212024_x_at	Hs.513984.1.1	-0.19939783	3.6893642	0.61814916	0.16754897	0
NIP30	224248_x_at	Hs.396740.1.5	-0.26410723	3.6844165	0.45900106	0.124579035	0
MED4	222438_at	Hs.181112.1.5	-0.47548842	3.630755	0.52520597	0.14465475	0
SF4	209547_s_at	Hs.515274.1.3	-0.056558482	3.6246328	0.45692265	0.1260604	0
ENOX2	234003_at	Hs.171458.1.30	1.8585044	3.5350497	0.31600183	0.08939106	0
TMEM50B	222907_x_at	Hs.433668.1.15	0.18478443	3.4882715	0.38473776	0.11029467	0
ACTL6A	202666_s_at	Hs.435326.1.30	0.29694545	3.4665406	0.39039993	0.11261946	0
ADRBK1	38447_at	Hs.83636.1.51	-0.7079206	3.4171197	0.38006258	0.11122308	0
POLR2C	216282_x_at	Hs.79402.1.17	-0.25801757	3.3467429	0.56638825	0.16923566	0
CLIP1	1558924_s_at	Hs.524809.1.31	-0.5087253	3.3453043	0.42631823	0.1274378	0
BCLAF1	214499_s_at	Hs.486542.1.16	0.69773984	3.307447	0.43667662	0.13202831	0
SNX13	1553148_a_at	Hs.487648.1.1	0.73240507	3.3070114	0.49322677	0.14914577	0
POLR2C	216282_x_at	Hs.79402.1.18	-0.2562784	3.2960167	0.5482681	0.16634263	0
RTN1	203485_at	Hs.368626.1.2	-0.4012655	3.291049	0.3278919	0.09963142	0
NNT	202783_at	Hs.482043.1.62	0.47200406	3.2693462	0.4403237	0.13468249	0
YARS2	218470_at	Hs.505231.1.4	-0.5885626	3.2283933	0.4579872	0.14186226	0
LMO7	202674_s_at	Hs.207631.1.53	-0.46177047	3.2051084	0.4192471	0.1308059	0
CREBBP	202160_at	Hs.459759.1.9	-0.2955316	3.1919832	0.45210016	0.14163613	0
GTPBP8	223486_at	Hs.127496.1.12	0.26339367	3.1885395	0.28116822	0.08818088	0
RB1	211540_s_at	Hs.408528.1.2	-0.47373608	3.1634471	0.2376648	0.07512843	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.7902869	3.1447191	0.6212666	0.1975587	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.7932024	3.1447191	0.6212666	0.1975587	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.7963622	3.1447191	0.6212666	0.1975587	0
UNKNOWN	235386_at	Hs.400256.1.1	0.32236934	3.1162353	0.25403038	0.08151836	0
UNKNOWN	235386_at	Hs.400256.1.2	0.32386726	3.1162353	0.25403038	0.08151836	0
ITGA11	222899_at	Hs.436416.1.2	-0.3223847	3.0826063	0.25055486	0.0812802	0
POLR2C	216282_x_at	Hs.79402.1.16	-0.25949863	3.069342	0.51216245	0.16686393	0
CTBS	218923_at	Hs.513557.1.5	-1.3045793	3.0410843	0.37340987	0.1227884	0
KLHL24	242088_at	Hs.407709.1.19	0.30032253	3.0387182	0.3623553	0.119246095	0
APIP	218698_at	Hs.447794.1.3	-0.7245199	2.988568	0.47715735	0.15966086	0
MON2	212754_s_at	Hs.389378.1.29	-0.55454624	2.9790182	0.28909642	0.09704419	0
LGR4	218326_s_at	Hs.502176.1.5	-0.7475958	2.9476888	0.20736842	0.07034949	0
HACE1	227471_at	Hs.434340.1.2	0.6630513	2.9372747	0.2970385	0.101127245	0
SLC48A1	48106_at	Hs.438867.1.13	-0.58228606	2.93263	0.37540615	0.12801006	0
IGL@	215121_x_at	Hs.449585.1.72	0.20729142	2.9080064	1.0033171	0.3450189	0
PHKB	202739_s_at	Hs.78060.1.71	-0.26996362	2.9063892	0.24666256	0.08486907	0
ZNF585A	243790_at	Hs.390568.1.6	-0.047627486	2.8710015	0.30698514	0.10692615	0
TMEM37	1554485_s_at	Hs.26216.1.6	0.05704582	2.8706043	0.22457668	0.07823324	0
CALCOCO2	235076_at	Hs.514920.1.38	-0.16717821	2.8570352	0.3462445	0.121190146	0
VPS4B	218171_at	Hs.126550.1.2	-0.09804824	2.848975	0.31514668	0.11061757	0
PNRC1	209034_at	Hs.75969.1.7	0.65846807	2.8453615	0.6042292	0.21235587	0
NIP30	223406_x_at	Hs.396740.1.6	-0.2608353	2.8153582	0.26222345	0.093140356	0
MOBP	207659_s_at	Hs.121333.1.11	0.24250226	2.7931275	0.21527743	0.07707397	0
SNX11	53912_at	Hs.15827.1.19	-0.17027523	2.792445	0.45898652	0.16436726	0
CSTF1	32723_at	Hs.172865.1.25	0.17556119	2.7501636	0.33117592	0.12042045	0
RPP30	1556061_at	Hs.139120.1.36	-0.84628946	2.7346492	0.16834205	0.06155892	0
SNHG10	238691_at	Hs.448753.1.2	-0.371089	2.733606	0.3623278	0.13254572	0
40067	201308_s_at	Hs.128199.1.37	0.37719503	2.7141738	0.34804696	0.12823312	0
ABHD4	242023_at	Hs.445665.1.22	-0.44237617	2.7093523	0.18536755	0.06841767	0
TIFA	238858_at	Hs.310640.1.1	0.3994023	2.6992638	0.34353366	0.12726939	0
FBXO8	223240_at	Hs.76917.1.3	0.43677327	2.695668	0.36842006	0.13667116	0
NEFL	221801_x_at	Hs.521461.1.2	0.8774212	2.694919	0.29652935	0.110032745	0
MYL4	217274_x_at	Hs.463300.1.13	-0.17448144	2.670489	0.25654054	0.096065	0
DCK	203302_at	Hs.709.1.27	0.36610824	2.6531246	0.26073003	0.098272815	0

Table B.4 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
MGAT2	203102_s_at	Hs.93338.1.10	-0.42782715	2.6483915	0.30349803	0.11459712	0
ADH5	208847_s_at	Hs.78989.1.10	0.39260113	2.6354413	0.25446987	0.096556835	0
NSDHL	215093_at	Hs.57698.1.21	2.194395	2.6285584	0.36804557	0.14001803	0
AGBL5	231857_s_at	Hs.138207.1.20	0.007687399	2.6282992	0.16093221	0.06123055	0
CREB1	204314_s_at	Hs.516646.1.28	0.110208705	2.6271262	0.22811323	0.086829945	0
CLDN22	222738_at	Hs.333179.1.38	0.4471084	2.6259105	0.27670875	0.1053763	0
SDF4	217855_x_at	Hs.42806.1.2	-2.313653	2.6093745	0.3171646	0.12154813	0
KLHL20	210634_at	Hs.495035.1.12	-1.1033049	2.6073773	0.23066545	0.088466465	0
CYP20A1	219565_at	Hs.446065.1.25	0.10742639	2.5981414	0.29329497	0.11288645	0
CYP20A1	219565_at	Hs.446065.1.24	0.10563422	2.5981414	0.29329497	0.11288645	0
TROVE2	212839_s_at	Hs.288178.1.26	-1.0312337	2.5898104	0.4029194	0.15557873	0
TDGF1	206286_s_at	Hs.385870.1.17	0.24538206	2.5853646	0.18523747	0.071648486	0
C10orf119	217905_at	Hs.124246.1.4	-0.81055474	2.5828035	0.36079562	0.13969147	0
LASS6	242019_at	Hs.506829.1.15	0.06964455	2.5745268	0.19697063	0.07650751	0
CCDC90B	222577_at	Hs.368866.1.1	-0.68432176	2.5704267	0.1875568	0.07296719	0
CCDC112	235208_at	Hs.436121.1.2	0.521852	2.5529895	0.39883536	0.15622288	0
HFE	211331_x_at	Hs.233325.1.1	0.6154833	2.5504506	0.16225459	0.06361801	0
LRRCC1	231872_at	Hs.193115.1.27	0.9348051	2.546739	0.31444103	0.12346809	0
CETN1	207209_at	Hs.122511.1.1	-0.12704207	2.5367012	0.17958885	0.07079622	0
ABHD4	242023_at	Hs.445665.1.23	-0.44050667	2.527464	0.17732057	0.070157506	0
XRN1	1555785_a_at	Hs.435103.1.4	0.28337267	2.5204794	0.2729519	0.108293645	0
CALCOCO2	235076_at	Hs.514920.1.37	-0.16890275	2.5024161	0.31237555	0.124829575	0
C21orf70	238703_at	Hs.410830.1.10	0.19981486	2.4926698	0.1771261	0.07105879	0
CEP76	52285_f_at	Hs.236940.1.5	-0.11687856	2.4906936	0.31909817	0.12811619	0
LGI4	242670_at	Hs.65256.1.20	-0.050624505	2.4870164	0.15769053	0.0634055	0
SSX2IP	203019_x_at	Hs.22587.1.2	-1.2876511	2.4832702	0.2613324	0.1052372	0
MCM10	222962_s_at	Hs.198363.1.34	-0.9361545	2.4808872	0.26956287	0.10865583	0
UNKNOWN	216304_x_at	Hs.499145.1.10	-0.91954935	2.4775493	0.5363455	0.21648227	0
ZNF764	222120_at	Hs.132227.1.3	-0.271615	2.4750068	0.32165325	0.12996055	0
TMEM33	218465_at	Hs.31082.1.28	0.3492443	2.4743276	0.18530336	0.07489039	0
LAS1L	235541_at	Hs.522675.1.13	1.3631772	2.4565427	0.20552811	0.083665594	0
ZNF839	219086_at	Hs.106005.1.22	-0.36454648	2.4535863	0.21115679	0.086060464	0
TFPI	209676_at	Hs.516578.1.9	0.09048575	2.4529078	0.3386798	0.13807277	0
TFPI	209676_at	Hs.516578.1.6	0.08869848	2.4529078	0.3386798	0.13807277	0
ATP7A	205198_s_at	Hs.496414.1.21	1.4220439	2.4512084	0.2989338	0.12195365	0
BCLAF1	201083_s_at	Hs.486542.1.12	0.6957755	2.4303048	0.19805771	0.08149501	0
TBCEL	231997_at	Hs.504136.1.20	-0.6275167	2.4221647	0.16197933	0.06687379	0
TET2	235461_at	Hs.367639.1.10	0.39769796	2.4134784	0.41825444	0.17329943	0
ACOX3	204241_at	Hs.479122.1.6	0.32723856	2.4083903	0.22971997	0.0953832	0
NIPBL	212483_at	Hs.481927.1.71	0.46813238	2.4039896	0.22817421	0.09491481	0
LCP2	244251_at	Hs.304475.1.1	0.57207596	2.3893535	0.20624292	0.08631746	0
GPR1	214605_x_at	Hs.184907.1.3	0.108920336	2.3788831	0.37032294	0.15567093	0
ERG	211626_x_at	Hs.473819.1.6	0.18916667	2.367823	0.30364585	0.12823841	0
EBNA1BP2	201323_at	Hs.346868.1.2	-1.4256233	2.3676736	0.25232375	0.106570326	0
AMOT	209521_s_at	Hs.528051.1.1	1.6151628	2.356813	0.400653	0.1699978	0
ESF1	218859_s_at	Hs.369284.1.19	0.14030258	2.3439126	0.28355068	0.12097323	0.2648
AK2	205996_s_at	Hs.470907.1.3	-1.561971	2.3397284	0.3279798	0.14017858	0.2648
QSER1	244563_at	Hs.369368.1.22	-0.7375305	2.3309097	0.30408084	0.13045587	0.2625
ZNF430	238614_x_at	Hs.466289.1.10	-0.055141203	2.3155913	0.2679122	0.11569926	0.5073
TBCC	202495_at	Hs.75064.1.2	0.63894165	2.3140612	0.2928212	0.12653996	0.5073
ANKRD11	231999_at	Hs.335003.1.7	-0.22512262	2.311085	0.33403164	0.14453456	0.5073
FDX1	203646_at	Hs.744.1.11	-0.6517416	2.3099246	0.45466614	0.1968316	0.5073
BBS10	219487_at	Hs.96322.1.3	-0.53216916	2.2946846	0.3073951	0.13395962	0.4947
TAF3	235119_at	Hs.527688.1.15	-0.95115083	2.2930324	0.30632424	0.13358915	0.4947
MAP3K7IP3	227357_at	Hs.188256.1.3	1.2763389	2.2892482	0.22740066	0.09933421	0.4947
DOPEY1	40612_at	Hs.520246.1.28	0.65629786	2.2832413	0.18819019	0.08242238	0.4907
ERG	211626_x_at	Hs.473819.1.5	0.18775691	2.253765	0.28699708	0.12734117	0.4828
RGS10	204316_at	Hs.501200.1.2	-0.8139964	2.2497902	0.41026366	0.1823564	0.4828
TMEM184C	219074_at	Hs.203896.1.18	0.42559958	2.2426178	0.2685094	0.11973034	0.4713
ZFP1	226807_at	Hs.388813.1.7	-0.23716433	2.239934	0.21928138	0.09789636	0.4713
REM1	210300_at	Hs.247729.1.9	0.15375659	2.2343495	0.16335793	0.07311208	0.4713

Table B.5 Significantly Shortened Genes in RAEB2 Subtype of MDS

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
SLC46A3	214719_at	Hs.117167.1.1	-0.47332403	4.105356	0.5088177	0.123939954	0
ADRBK1	38447_at	Hs.83636.1.51	-0.6900677	3.7684062	0.4021926	0.1067275	0
SF4	209547_s_at	Hs.515274.1.3	-0.043354146	3.6419587	0.4668373	0.12818302	0
STAC	205743_at	Hs.56045.1.18	0.24529886	3.5908225	0.3374209	0.09396759	0
LGI4	242670_at	Hs.65256.1.20	-0.037803017	3.588495	0.22992358	0.06407242	0
TET2	235461_at	Hs.367639.1.10	0.40311038	3.4599924	0.5872738	0.16973267	0
ENOX2	234003_at	Hs.171458.1.30	1.8582339	3.4266648	0.30914772	0.09021825	0
ZAK	218833_at	Hs.444451.1.34	0.08649528	3.381661	0.47634768	0.14086205	0
PNRC1	209034_at	Hs.75969.1.7	0.6614158	3.3703701	0.7046807	0.2090811	0
ACTL6A	202666_s_at	Hs.435326.1.30	0.30427995	3.3642015	0.37792277	0.112336546	0
UNKNOWN	235386_at	Hs.400256.1.2	0.329099	3.3385918	0.24908516	0.07460785	0
UNKNOWN	235386_at	Hs.400256.1.1	0.32739207	3.3385918	0.24908516	0.07460785	0
LAS1L	235541_at	Hs.522675.1.13	1.3619603	3.3375552	0.31917772	0.095632195	0
VPS4B	218171_at	Hs.126550.1.2	-0.08664023	3.2802477	0.3622582	0.11043623	0
FLII	212024_x_at	Hs.513984.1.1	-0.18650216	3.2771263	0.53402495	0.16295525	0
MON2	212754_s_at	Hs.389378.1.29	-0.53793573	3.241521	0.30585843	0.09435646	0
SNX13	1553148_a_at	Hs.487648.1.1	0.7347679	3.1648011	0.48023152	0.15174146	0
RB1	211540_s_at	Hs.408528.1.2	-0.45833132	3.1202345	0.24759804	0.07935238	0
NIP30	223406_x_at	Hs.396740.1.6	-0.24575037	3.0605586	0.2947688	0.0963121	0
ACTR3	213102_at	Hs.433512.1.22	0.06771956	3.0373924	0.36564767	0.1203821	0
BCLAF1	201083_s_at	Hs.486542.1.12	0.6962699	3.0372245	0.23661762	0.07790587	0
POLR2C	216282_x_at	Hs.79402.1.18	-0.24133019	3.0278578	0.50528216	0.16687778	0
SPOCK3	235342_at	Hs.481133.1.3	0.43706208	2.9925616	0.35115728	0.11734337	0
ZFP1	226807_at	Hs.388813.1.7	-0.22319455	2.9696298	0.2949862	0.09933433	0
NEFL	221801_x_at	Hs.521461.1.2	0.8771158	2.9525552	0.33311558	0.11282281	0
POLR2C	216282_x_at	Hs.79402.1.17	-0.24285972	2.9412901	0.4983667	0.16943814	0
VASN	225867_at	Hs.372579.1.7	-0.2753217	2.930466	0.17944467	0.061234176	0
KLHL20	210634_at	Hs.495035.1.12	-1.0995975	2.9217472	0.23649317	0.08094238	0
BCLAF1	214499_s_at	Hs.486542.1.16	0.69866	2.8456004	0.4051423	0.14237498	0
CREBBP	202160_at	Hs.459759.1.9	-0.27863118	2.8309045	0.40426636	0.14280467	0
GTPBP8	223486_at	Hs.127496.1.12	0.26797566	2.777342	0.24970907	0.08990937	0
AGBL5	231857_s_at	Hs.138207.1.20	0.018546032	2.7720797	0.17207918	0.06207584	0
MAP3K7IP3	227357_at	Hs.188256.1.3	1.2764039	2.769702	0.2801783	0.10115829	0
ARL4A	205020_s_at	Hs.245540.1.6	0.72996706	2.7655213	0.27440944	0.099225216	0
ZKSCAN5	203731_s_at	Hs.110839.1.21	0.7910685	2.742466	0.21416402	0.07809177	0
C9orf82	231995_at	Hs.178357.1.3	1.055833	2.7302785	0.24728045	0.09056968	0
NNT	202783_at	Hs.482043.1.62	0.47198918	2.687335	0.39649796	0.14754318	0
POLR2C	216282_x_at	Hs.79402.1.16	-0.24433169	2.683503	0.4509604	0.16804916	0
C12orf4	218374_s_at	Hs.302977.1.4	-0.5980716	2.6720793	0.20287794	0.07592512	0
TMEM33	218465_at	Hs.31082.1.28	0.35365397	2.6540682	0.20597547	0.07760745	0
CSE1L	201111_at	Hs.90073.1.50	0.17934355	2.650695	0.26511705	0.100017935	0
TNFSF10	214329_x_at	Hs.478275.1.2	0.30117163	2.6265152	0.40873957	0.15562049	0
SCARB2	201647_s_at	Hs.349656.1.9	0.3784919	2.623014	0.20516425	0.078216985	0
YARS2	218470_at	Hs.505231.1.4	-0.57314193	2.616476	0.37432826	0.14306581	0
CREB1	204314_s_at	Hs.516646.1.28	0.12368398	2.6121004	0.25353342	0.097061135	0
NIPBL	212483_at	Hs.481927.1.71	0.46851474	2.5795171	0.24744558	0.0959271	0
ZNF554	242864_at	Hs.307043.1.14	-0.06866047	2.5525718	0.22661704	0.08877989	0
FAM5B	214822_at	Hs.495918.1.7	-1.089079	2.5519516	0.13030353	0.051060345	0
HACE1	227471_at	Hs.434340.1.2	0.6656783	2.5089965	0.24000743	0.095658734	0
AGBL5	231857_s_at	Hs.138207.1.21	0.020168306	2.5071805	0.17153916	0.06841915	0
NIN	224304_x_at	Hs.310429.1.13	-0.40176642	2.4998617	0.17346656	0.06939046	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.786164	2.4978898	0.5203464	0.2083144	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.7798975	2.4978898	0.5203464	0.2083144	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.78328633	2.4978898	0.5203464	0.2083144	0
CCNB1	228729_at	Hs.23960.1.22	0.4841738	2.493332	0.2532642	0.101576604	0
FGB	204988_at	Hs.300774.1.15	0.43340883	2.4851124	0.18475889	0.07434629	0
NIP30	224248_x_at	Hs.396740.1.5	-0.24866544	2.4793932	0.31918788	0.12873629	0
KIAA1430	226254_s_at	Hs.535734.1.4	0.45126578	2.4755504	0.25719133	0.10389259	0
DCK	203302_at	Hs.709.1.27	0.37009257	2.4718144	0.26752186	0.108228944	0
C9orf85	238579_at	Hs.534190.1.18	1.0745903	2.465814	0.25833255	0.104765624	0
ZNF625	244406_at	Hs.512823.1.2	-0.05189508	2.4514573	0.33800614	0.13787968	0
ZNF625	244406_at	Hs.512823.1.1	-0.053261705	2.4514573	0.33800614	0.13787968	0
CTBS	218923_at	Hs.513557.1.5	-1.3065157	2.4371052	0.29431856	0.120765634	0
C10orf119	217905_at	Hs.124246.1.4	-0.79986775	2.4355147	0.33893383	0.13916312	0
JRKL	206734_at	Hs.105940.1.2	-0.65454245	2.4353356	0.23337835	0.09583005	0
MED14	202610_s_at	Hs.407604.1.7	1.2901051	2.431256	0.23231423	0.095553175	0

Table B.5 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
CLIP1	1558924_s_at	Hs.524809.1.31	-0.49343172	2.4113705	0.34183043	0.14175774	0
EBNA1BP2	201323_at	Hs.346868.1.2	-1.4411404	2.3871014	0.28937358	0.12122383	0
LYPLA2	215568_x_at	Hs.533479.1.18	-1.8064655	2.3794196	0.29308677	0.12317574	0
RAB18	224377_s_at	Hs.406799.1.9	-0.9043765	2.3786886	0.22644514	0.09519747	0
CUL4B	215997_s_at	Hs.102914.1.20	1.7354878	2.3761947	0.1295795	0.054532357	0
MAP2K5	211370_s_at	Hs.114198.1.23	-0.3061624	2.3759046	0.20823777	0.08764568	0
TPRG1	229764_at	Hs.338851.1.15	0.31068313	2.3585224	0.17039692	0.07224732	0
ZNF143	221873_at	Hs.523471.1.34	-0.73962754	2.3537457	0.20441943	0.086848564	0
CREM	214508_x_at	Hs.200250.1.29	-0.8934325	2.3491535	0.19185793	0.08167109	0
MGAT2	203102_s_at	Hs.93338.1.10	-0.41186816	2.3488064	0.27334893	0.1163778	0
PCYT1A	204210_s_at	Hs.435767.1.20	0.3158411	2.333135	0.17554826	0.075241365	0
PCYT1A	204210_s_at	Hs.435767.1.19	0.3144405	2.333135	0.17554826	0.075241365	0
ANKRD11	231999_at	Hs.335003.1.7	-0.21200997	2.3313441	0.32567966	0.13969609	0
CCNB1	228729_at	Hs.23960.1.20	0.48250416	2.3237886	0.20560828	0.088479765	0
ADSL	210250_x_at	Hs.75527.1.19	0.23603931	2.304041	0.29644454	0.12866288	0.8917
TFPI2	209277_at	Hs.438231.1.11	0.77504593	2.2855206	0.1706214	0.074653186	0.8702
TPT1	214327_x_at	Hs.374596.1.15	-0.46382907	2.283615	0.21696894	0.09501117	0.8702
MEF2A	208328_s_at	Hs.268675.1.28	-0.28512245	2.2683384	0.30847144	0.13599004	0.8497
SMARCA4	208794_s_at	Hs.327527.1.81	-0.05895018	2.2585752	0.1966781	0.087080605	0.8399
LARS	222428_s_at	Hs.432674.1.6	0.5515365	2.253346	0.35232747	0.15635747	0.8302
ZNF440	241731_x_at	Hs.418192.1.5	-0.056208443	2.212703	0.15985635	0.07224483	0.8025
SNHG10	238691_at	Hs.448753.1.2	-0.35317382	2.2097487	0.29737604	0.13457459	0.8025
LOC440131	233804_at	Hs.132371.1.1	-0.4716823	2.2088192	0.11656834	0.052774053	0.8025
IL1RAP	210233_at	Hs.478673.1.18	0.31267202	2.190034	0.2766707	0.12633169	0.7937
ZNF839	219086_at	Hs.106005.1.22	-0.3461281	2.1818342	0.18940395	0.08680951	0.7851
POT1	204354_at	Hs.31968.1.4	0.8190685	2.1736085	0.18238473	0.08390873	0.7766
ZNF585A	243790_at	Hs.390568.1.6	-0.03536978	2.163743	0.26329994	0.12168725	1.1404
TDGF1	206286_s_at	Hs.385870.1.17	0.25027725	2.1610508	0.15084198	0.069800295	1.1404
BCAR3	204032_at	Hs.36958.1.4	-1.2643046	2.1505723	0.16522425	0.07682804	1.1286
CCNG2	202769_at	Hs.13291.1.25	0.3845324	2.1170044	0.3271981	0.15455711	1.4741
ITGA11	222899_at	Hs.436416.1.2	-0.3046332	2.1138864	0.17757148	0.084002376	1.4741
MOBP	207659_s_at	Hs.121333.1.11	0.2470321	2.0977876	0.15534455	0.07405161	1.4592
XRN1	1555785_a_at	Hs.435103.1.4	0.28948683	2.0883327	0.21170142	0.10137342	1.4446
CALCOCO2	235076_at	Hs.514920.1.38	-0.15369982	2.081985	0.24822742	0.11922632	1.4303
FDPS	201275_at	Hs.335918.1.24	-1.1424954	2.056703	0.21711487	0.10556451	1.4163
COG1	231813_s_at	Hs.283109.1.40	-0.12372994	2.0511885	0.20256513	0.09875501	1.3758
CCL20	205476_at	Hs.75498.1.4	0.12977785	2.0455844	0.12667358	0.061925374	1.3758
COL11A2	213870_at	Hs.390171.1.2	0.62446576	2.0435789	0.102832444	0.050319783	1.3758
EMR2	232009_at	Hs.531619.1.3	-0.047700383	2.0321238	0.36247838	0.17837416	1.3376
PROC	206259_at	Hs.224698.1.14	0.0705435	2.030537	0.117495105	0.05786406	1.3376
IQCK	215131_at	Hs.460217.1.5	-0.268903	2.0275774	0.4961996	0.24472536	1.3376
EMR2	232009_at	Hs.531619.1.2	-0.04915836	2.0008626	0.3582635	0.17905451	1.9347
DYNC1I1	205348_s_at	Hs.440364.1.34	0.77776265	1.9969379	0.13185889	0.06603054	1.9347
TMEM182	1555090_x_at	Hs.436203.1.10	0.06329934	1.9966124	0.17136869	0.08582972	1.9347
SCARB2	201647_s_at	Hs.349656.1.10	0.37989452	1.9934328	0.17376563	0.08716904	1.9347
GAS7	211067_s_at	Hs.462214.1.9	-0.19045419	1.9858801	0.18240514	0.09185103	1.9008
SRP72	208800_at	Hs.237825.1.32	0.36513275	1.9840076	0.1719222	0.08665401	1.9008
KLHL24	242088_at	Hs.407709.1.19	0.30751833	1.9702879	0.23683798	0.12020476	1.8842
RMND5A	212479_s_at	Hs.75277.1.19	0.04945761	1.957064	0.19833282	0.101342015	1.8209
ATP7A	205198_s_at	Hs.496414.1.21	1.4490671	1.9552938	0.239308	0.12238979	1.8209
RTN1	203485_at	Hs.368626.1.2	-0.3843337	1.9528102	0.20412949	0.10453115	1.8209
PPFIA1	202066_at	Hs.530749.1.60	-0.68290216	1.9517689	0.14455435	0.07406325	1.8209

Table B.6 Significantly Lengthened Gene in RAEB2 Subtype of MDS

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
PABPC3	208113_x_at	Hs.458280.1.1	-0.4765717	-3.5345535	-0.7465235	0.2112073	0.8598977

Table B.7 Significantly Shortened Genes in Glioblastoma - GSE4290 dataset

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (r)	q-value
KIAA1245	214693_x_at	Hs.515947.1.33	-1.2851444	9.155828	0.8454592	0.092341095	0
SNX3	213545_x_at	Hs.12102.1.3	0.6707688	8.64416	0.47684944	0.055164345	0
RPL13	212933_x_at	Hs.410817.1.11	-0.24764684	8.325072	0.7309533	0.087801434	0
APP	200602_at	Hs.434980.1.2	0.18803586	7.846167	0.456918	0.05823455	0
SNX3	200067_x_at	Hs.12102.1.3	0.66733223	7.4034653	0.48213303	0.06512262	0
ESF1	218859_s_at	Hs.369284.1.19	0.14378928	7.182947	0.35635108	0.049610704	0
KRIT1	204738_s_at	Hs.531987.1.8	0.81615925	7.1380644	0.43594465	0.06107323	0
VASP	202205_at	Hs.515469.1.28	-0.043611098	6.823919	0.6324452	0.09268065	0
CBFA2T3	208056_s_at	Hs.513811.1.2	-0.24982186	6.123053	0.47419047	0.07744347	0
ATP7A	205198_s_at	Hs.496414.1.21	1.5047401	6.0661893	0.2564004	0.04226713	0
TTC30A	213679_at	Hs.128384.1.2	0.07734493	6.033815	0.44941118	0.0744821	0
RLF	204243_at	Hs.205627.1.5	-1.5831281	5.673151	0.26847312	0.047323458	0
CHAF1A	214426_x_at	Hs.79018.1.40	-0.0886061	5.6248627	0.3329551	0.059193462	0
BGN	201262_s_at	Hs.821.1.15	2.0957444	5.585268	0.6636007	0.11881268	0
CARTPT	206339_at	Hs.1707.1.4	0.50821096	5.5509534	0.5353747	0.09644734	0
ZAK	218833_at	Hs.444451.1.34	0.0736021	5.377546	0.47812384	0.08891116	0
SLC16A1	209900_s_at	Hs.75231.1.3	-1.2933382	5.317038	0.31458694	0.059165824	0
SRD5A1	211056_s_at	Hs.552.1.20	0.47668457	5.074107	0.42591086	0.08393809	0
SRD5A1	211056_s_at	Hs.552.1.19	0.4738349	5.074107	0.42591086	0.08393809	0
TNPO1	209225_x_at	Hs.482497.1.48	0.51273125	5.0273747	0.38359168	0.07630059	0
KCNN1	206231_at	Hs.158173.1.9	-0.06492406	4.9009333	0.28531417	0.058216292	0
TCF3	213730_x_at	Hs.371282.1.8	-0.102044806	4.8223133	0.33042008	0.068518996	0
TOP2A	201291_s_at	Hs.156346.1.29	-0.2067253	4.82178	0.43003917	0.0891868	0
TNPO1	209225_x_at	Hs.482497.1.46	0.51074475	4.811168	0.58132	0.1208272	0
DDOST	208674_x_at	Hs.523145.1.8	-2.018588	4.8029017	0.24571362	0.05115941	0
CBX7	212914_at	Hs.356416.1.3	0.22830562	4.7763343	0.3524246	0.07378558	0
STMN1	200783_s_at	Hs.209983.1.5	-1.818357	4.7129855	0.27385163	0.058105767	0
DNPEP	38703_at	Hs.258551.1.5	0.121805586	4.690095	0.33442765	0.071305096	0
TMEM164	220486_x_at	Hs.496572.1.20	1.6297799	4.6426773	0.20865819	0.044943504	0
SPINT2	210715_s_at	Hs.31439.1.18	-0.05392342	4.6330843	0.23551945	0.05083427	0
THBS2	203083_at	Hs.371147.1.3	0.73399633	4.5678854	0.22374478	0.048982136	0
XYLT1	213725_x_at	Hs.22907.1.3	-0.31725144	4.530335	0.35953394	0.079361446	0
HNRNPR	208766_s_at	Hs.373763.1.5	-1.9283783	4.4168477	0.15147398	0.03429459	0
PCM1	214937_x_at	Hs.491148.1.105	0.9065509	4.361985	0.24446298	0.05604397	0
USP12	215886_x_at	Hs.42400.1.21	-0.56863254	4.3410387	0.35128134	0.08092103	0
SKP1	200711_s_at	Hs.171626.1.6	0.5501366	4.284412	0.17098105	0.039907705	0
CYP4A11	207407_x_at	Hs.1645.1.4	-1.4950768	4.2733016	0.24861065	0.05817765	0
AGTPBP1	204500_s_at	Hs.494321.1.5	1.1060438	4.2599072	0.22708157	0.05330669	0
STK38	202951_at	Hs.409578.1.2	0.62101156	4.188756	0.17812854	0.042525403	0
ZNF207	200829_x_at	Hs.500775.1.32	-0.22132508	4.168074	0.1416882	0.033993684	0
SETBP1	205933_at	Hs.435458.1.6	-0.12912253	4.1185637	0.26921833	0.06536704	0
MGAT4A	219797_at	Hs.177576.1.19	0.026158564	4.0873694	0.17377737	0.0425157	0
CAPZB	201949_x_at	Hs.432760.1.4	-2.0887253	4.070497	0.18756604	0.046079393	0
WIPI1	203827_at	Hs.463964.1.12	-0.1611721	4.022955	0.1944505	0.048335243	0
RPS6KB1	204171_at	Hs.463642.1.34	-0.17030004	4.0128474	0.23941335	0.059661712	0
SLC16A3	217691_x_at	Hs.500761.1.20	-0.15652983	4.010677	0.4930116	0.12292478	0
RNASET2	217984_at	Hs.529989.1.9	0.7304351	3.9933276	0.18911833	0.04735858	0
CDC42EP3	209288_s_at	Hs.369574.1.3	2.65E-04	3.9633343	0.14125516	0.035640486	0
CDC42EP4	218063_s_at	Hs.3903.1.8	-0.15929276	3.922882	0.1288375	0.03284256	0
SPIN1	217813_s_at	Hs.146804.1.15	1.1124301	3.9173985	0.16154094	0.04123679	0
DAZAP2	214334_x_at	Hs.369761.1.9	-0.6584349	3.9040697	0.32812202	0.08404615	0
SKP1	200711_s_at	Hs.171626.1.7	0.5526907	3.862673	0.13570717	0.035132967	0
AURKA	204092_s_at	Hs.250822.1.2	0.1730591	3.859411	0.2776314	0.07193621	0
ARF6	203312_x_at	Hs.525330.1.12	-0.49980474	3.8263664	0.28104252	0.073448926	0
INPP5A	203006_at	Hs.523360.1.34	-0.88249844	3.8036094	0.13202295	0.03470991	0
PMPCB	201682_at	Hs.184211.1.24	0.8622471	3.7614229	0.14262009	0.037916526	0
RPL12	214271_x_at	Hs.408054.1.12	1.2014039	3.7571783	0.10904151	0.029022181	0
PTPRM	203329_at	Hs.49774.1.58	-0.14578287	3.7205825	0.21924515	0.05892764	0
FAM82B	218549_s_at	Hs.145386.1.7	0.9893622	3.7060816	0.16973281	0.045798454	0

Table B.8 Significantly Lengthened Genes in Glioblastoma - GSE4290 dataset

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (r)	q-value
KPNA1	202057_at	Hs.161008.1.1	0.28225282	-10.103333	-0.5425848	0.05370354	0
DLGAP5	203764_at	Hs.77695.1.6	-0.48756215	-7.3392243	-0.45401254	0.061861105	0
LOC157627	214839_at	Hs.12513.1.2	0.89500964	-7.081796	-0.49687755	0.07016265	0
SSX2IP	203019_x_at	Hs.22587.1.2	-1.3592836	-7.0650206	-0.6612082	0.093589	0
PSMC6	201699_at	Hs.156171.1.27	-0.4921745	-6.996092	-0.70178264	0.10031067	0
PSMC6	201699_at	Hs.156171.1.26	-0.49459016	-6.996092	-0.70178264	0.10031067	0
EIF5B	201024_x_at	Hs.158688.1.13	0.03045038	-6.9504457	-0.57111967	0.082170226	0
RAD17	207405_s_at	Hs.16184.1.34	0.50519216	-6.7306848	-0.26291582	0.03906227	0
SF3A2	209381_x_at	Hs.115232.1.23	-0.09538068	-6.321973	-0.5075875	0.080289416	0
RAB13	202252_at	Hs.151536.1.5	-1.2529346	-6.2601905	-0.7630412	0.121887855	0
SCCPDH	201826_s_at	Hs.498397.1.19	-1.0427165	-6.22216	-0.32279158	0.051877737	0
BUB1	209642_at	Hs.469649.1.11	0.041582853	-6.1948457	-0.3305217	0.053354308	0
NMT2	215743_at	Hs.60339.1.1	-0.99118507	-6.109695	-0.31860483	0.05214742	0
GSTM5	205752_s_at	Hs.75652.1.12	-1.3113345	-6.094957	-0.539512	0.08851777	0
GNS	203676_at	Hs.334534.1.5	-0.6354814	-6.0281477	-0.27723277	0.04598971	0
SCCPDH	201826_s_at	Hs.498397.1.20	-1.0375973	-5.834925	-0.36387134	0.062360927	0
TBX2	40560_at	Hs.531085.1.20	-0.16797443	-5.792545	-0.48831093	0.0842999	0
CLU	208792_s_at	Hs.436657.1.10	0.9390269	-5.730942	-0.24286658	0.042378128	0
PHTF2	215286_s_at	Hs.203965.1.18	0.80670726	-5.635139	-0.2975359	0.0528001	0
PHTF2	215286_s_at	Hs.203965.1.19	0.809845	-5.635139	-0.2975359	0.0528001	0
TFDP1	212330_at	Hs.79353.1.30	-0.51247394	-5.424301	-0.32683605	0.060254037	0
PBX2	202875_s_at	Hs.509545.1.5	0.6122947	-5.3214316	-0.28577304	0.053702284	0
KYNU	210663_s_at	Hs.470126.1.17	0.054101933	-5.307425	-0.4111265	0.07746252	0
KYNU	210663_s_at	Hs.470126.1.16	0.05219946	-5.307425	-0.4111265	0.07746252	0
KYNU	210663_s_at	Hs.470126.1.18	0.05604765	-5.307425	-0.4111265	0.07746252	0
HMOX2	218121_at	Hs.284279.1.14	-0.32189107	-5.2257323	-0.38518655	0.073709585	0
KCNMB4	219287_at	Hs.525529.1.8	-0.6251377	-5.183068	-0.47563535	0.09176715	0
CCL11	210133_at	Hs.54460.1.3	-0.21725164	-5.104347	-0.22044167	0.043187045	0
TMEM184C	219074_at	Hs.203896.1.17	0.4399405	-5.0309787	-0.34171557	0.06792229	0
RB1	211540_s_at	Hs.408528.1.2	-0.5364557	-4.9556627	-0.2666604	0.05380923	0
PTK2	207821_s_at	Hs.395482.1.8	1.0434622	-4.9094214	-0.30464536	0.06205321	0
BMP7	211259_s_at	Hs.473163.1.16	0.17934138	-4.902223	-0.20676789	0.042178392	0
BMP7	211259_s_at	Hs.473163.1.17	0.18112849	-4.902223	-0.20676789	0.042178392	0
CUGBP2	202158_s_at	Hs.309288.1.43	-1.013292	-4.898262	-0.38741696	0.07909274	0
LETM1	218939_at	Hs.120165.1.4	0.35073796	-4.8245792	-0.23737375	0.049200922	0
PGLYRP4	220944_at	Hs.58356.1.2	-1.2615241	-4.8049655	-0.26385444	0.05491287	0
DDRGK1	218159_at	Hs.471975.1.3	0.1394699	-4.8046064	-0.31854528	0.066299975	0
WDTC1	40829_at	Hs.469154.1.38	-1.7840714	-4.768916	-0.29923797	0.06274758	0
CD44	210916_s_at	Hs.502328.1.28	-0.80946743	-4.75793	-0.27364635	0.057513744	0
CD44	210916_s_at	Hs.502328.1.11	-0.8164272	-4.75793	-0.27364635	0.057513744	0
CD44	210916_s_at	Hs.502328.1.13	-0.81311154	-4.75793	-0.27364635	0.057513744	0
IQGAP1	213446_s_at	Hs.430551.1.35	-0.34886426	-4.756597	-0.5490771	0.115434855	0
DOPEY1	40612_at	Hs.520246.1.28	0.65233994	-4.56494	-0.3075388	0.06736974	0
TBCC	202495_at	Hs.75064.1.2	0.62938726	-4.544272	-0.29860675	0.06571058	0
AMOT	209521_s_at	Hs.528051.1.1	1.651872	-4.544024	-0.6169982	0.13578233	0
POSTN	210809_s_at	Hs.136348.1.11	-0.5475284	-4.4788756	-0.6967979	0.15557429	0
COL4A1	211980_at	Hs.17441.1.3	-0.52597266	-4.4095025	-0.2858488	0.06482563	0
LMO7	202674_s_at	Hs.207631.1.53	-0.5338292	-4.365475	-0.352495	0.080746084	0
MTDH	212250_at	Hs.377155.1.34	1.0186074	-4.2979364	-0.18319082	0.042622972	0
TSKU	218245_at	Hs.8361.1.6	-0.7687829	-4.2472205	-0.29711643	0.0699555	0
MKLN1	204423_at	Hs.44693.1.21	0.88042754	-4.2426085	-0.22456947	0.05293193	0
TEX261	212084_at	Hs.516087.1.4	0.013441303	-4.2389174	-0.21229821	0.050083123	0
QPCT	205174_s_at	Hs.79033.1.18	-0.001649671	-4.2126837	-0.29925764	0.07103729	0
SLC48A1	48106_at	Hs.438867.1.13	-0.66181386	-4.142256	-0.35335505	0.08530498	0
TLE4	214688_at	Hs.444213.1.41	1.0934141	-4.1142955	-0.38535392	0.09366219	0
NFIC	213298_at	Hs.170131.1.21	-0.09309571	-4.1129746	-0.18972056	0.046127334	0
SLC45A2	220245_at	Hs.278962.1.3	0.48518074	-4.0988774	-0.15792346	0.038528465	0
MTMR2	214649_s_at	Hs.181326.1.8	-0.7498944	-4.0591145	-0.28121784	0.06928059	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.89882094	-4.0512257	-0.3187275	0.07867434	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.89495665	-4.0512257	-0.3187275	0.07867434	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.8908697	-4.0512257	-0.3187275	0.07867434	0
NEIL3	219502_at	Hs.405467.1.23	0.46597138	-4.050965	-0.23117863	0.057067547	0
RNF10	208632_at	Hs.442798.1.40	-0.588449	-4.0229025	-0.17587747	0.04371905	0
SFRS6	206108_s_at	Hs.6891.1.21	0.16473842	-3.971723	-0.23173231	0.058345534	0
TFDP1	212330_at	Hs.79353.1.29	-0.5152261	-3.9411824	-0.17706645	0.04492724	0
CSDA	201160_s_at	Hs.221889.1.4	-0.6839145	-3.8627605	-0.25511885	0.06604573	0
MFAP3	214588_s_at	Hs.432818.1.12	0.5718564	-3.8251011	-0.16861904	0.044082243	0

Table B.8 (continued)

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (r)	q-value
TFPI	209676_at	Hs.516578.1.9	0.08638941	-3.820635	-0.30165315	0.07895367	0
TFPI	209676_at	Hs.516578.1.6	0.083955586	-3.820635	-0.30165315	0.07895367	0
TRA2A	204658_at	Hs.445652.1.7	0.7717036	-3.815709	-0.27802154	0.07286235	0
SMAD6	207069_s_at	Hs.153863.1.12	-0.3802904	-3.8105612	-0.28693098	0.075298876	0
SMAD6	207069_s_at	Hs.153863.1.11	-0.38270667	-3.8105612	-0.28693098	0.075298876	0
TSKU	218245_at	Hs.8361.1.5	-0.7719133	-3.7986462	-0.2898806	0.07631156	0
SMAD7	204790_at	Hs.465087.1.2	-0.12483605	-3.773076	-0.19016594	0.05040077	0
PPIC	204517_at	Hs.110364.1.4	0.54210705	-3.7471445	-0.30103695	0.080337696	0
MLPH	218211_s_at	Hs.102406.1.41	0.13031854	-3.6782947	-0.2620058	0.07123024	0
ARL4A	205020_s_at	Hs.245540.1.4	0.75340116	-3.6677399	-0.22812897	0.062198788	0
PTPN11	212610_at	Hs.506852.1.25	-0.5912818	-3.6249697	-0.15394816	0.042468812	0
AGPAT3	219723_x_at	Hs.248785.1.30	0.20028673	-3.6156864	-0.22836667	0.06315998	0
TMEM127	219460_s_at	Hs.355708.1.2	0.021585075	-3.6108704	-0.18151616	0.050269365	0
TMEM127	219460_s_at	Hs.355708.1.3	0.02403782	-3.6108704	-0.18151616	0.050269365	0
WAS	38964_r_at	Hs.2157.1.19	1.3830318	-3.6052587	-0.25307214	0.07019528	0
SLC15A2	205317_s_at	Hs.518089.1.16	0.27940103	-3.590505	-0.34893352	0.097182296	0

Table B.9 Significantly Shortened Genes in Glioblastoma - GSE16011 dataset

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (r)	q-value
TCF3	213730_x_at	Hs.371282.1.8	-0.045117836	7.5910277	1.1251845	0.14822559	0
METTL9	217868_s_at	Hs.279583.1.14	-0.23301116	7.524576	1.1247756	0.14948027	0
DNPEP	38703_at	Hs.258551.1.5	0.14001124	6.419508	0.5647367	0.08797196	0
ZAK	218833_at	Hs.444451.1.34	0.100508265	5.902108	0.8924185	0.15120335	0
METTL9	217868_s_at	Hs.279583.1.15	-0.23056412	5.7731957	0.8757744	0.15169664	0
SNX3	213545_x_at	Hs.12102.1.3	0.66934997	5.4117165	0.50596714	0.09349476	0
TCF3	213730_x_at	Hs.371282.1.10	-0.042393263	5.3239803	1.07851	0.20257589	0
TOP2A	201291_s_at	Hs.156346.1.29	-0.1420459	5.27135	0.6609243	0.12538047	0
TROVE2	212839_s_at	Hs.288178.1.26	-0.9512875	5.214033	0.687896	0.13193165	0
SNX3	200067_x_at	Hs.12102.1.3	0.6656148	5.1520963	0.47830343	0.09283666	0
RPL13	212933_x_at	Hs.410817.1.11	-0.18881972	5.1060266	0.64649016	0.12661316	0
SLC16A3	217691_x_at	Hs.500761.1.20	-0.089939885	5.062584	1.2951744	0.25583267	0
AP2A2	215764_x_at	Hs.19121.1.49	-0.7642485	4.910312	0.6935053	0.14123446	0
AURKA	204092_s_at	Hs.250822.1.2	0.18241991	4.8580427	0.5485177	0.1129092	0
MGAT2	203102_s_at	Hs.93338.1.10	-0.39122057	4.8377037	0.44101155	0.09116134	0
SEC22A	218703_at	Hs.477361.1.19	0.28455535	4.702643	0.44815502	0.09529854	0
ADH5	208847_s_at	Hs.78989.1.10	0.40722227	4.5483966	0.49548626	0.10893647	0
ADSL	210250_x_at	Hs.75527.1.19	0.24107091	4.314364	0.48007536	0.11127373	0
NAP1L1	204528_s_at	Hs.524599.1.9	-0.51137763	4.2849445	0.47113234	0.109950624	0
HRH1	205579_at	Hs.1570.1.7	0.25413612	4.2024193	0.39201635	0.0932835	0
COL1A2	202404_s_at	Hs.489142.1.70	0.8027565	4.1676693	0.5325861	0.12778991	0
DKC1	216212_s_at	Hs.4747.1.24	2.0970633	4.1641994	0.31883848	0.07656658	0
ACOX3	204241_at	Hs.479122.1.6	0.34621793	4.10903	0.32966834	0.08023021	0
CTBS	218923_at	Hs.513557.1.5	-1.2304376	4.067413	0.37518907	0.09224268	0
ETAA1	219216_at	Hs.353022.1.14	0.048430186	4.051974	0.49185598	0.12138676	0
MGAT4A	219797_at	Hs.177576.1.19	0.06538454	4.017731	0.45576683	0.11343885	0
CCDC93	219774_at	Hs.107845.1.3	0.08278598	3.9741771	0.33159855	0.08343829	0

Table B.10 Significantly Lengthened Genes in Glioblastoma - GSE4290 dataset

Gene Name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (r)	q-value
NEFH	204412_s_at	Hs.198760.1.10	0.22751209	-10.211063	-1.8569276	0.18185449	0
SSX2IP	203019_x_at	Hs.22587.1.2	-1.2107042	-7.0036488	-1.0785599	0.15399972	0
NMT2	215743_at	Hs.60339.1.1	-0.8860967	-6.850912	-0.5687601	0.083019614	0
LOC157627	214839_at	Hs.12513.1.2	0.8891248	-6.171595	-0.5383291	0.087226905	0
KIF5A	205318_at	Hs.151219.1.36	-0.537454	-5.65766	-0.48407772	0.08556147	0
SECISBP2L	212450_at	Hs.9997.1.3	-0.31712553	-5.543716	-0.6629315	0.11958252	0
VIP	206577_at	Hs.53973.1.18	0.70393336	-5.2577643	-0.42768258	0.08134305	0
TBX2	40560_at	Hs.531085.1.21	-0.0990635	-4.8558893	-0.46242222	0.09522915	0
SLC48A1	48106_at	Hs.438867.1.13	-0.5661179	-4.531198	-0.7401202	0.16333874	0

Table B.11 Significantly Shortened Genes in Gastric Cardia Adenocarcinoma

Gene name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
ATP6V1C1	202872_at	Hs.86905.1.21	1.0417829	9.345377	0.58879626	0.06300402	0
TIMM9	218316_at	Hs.440525.1.4	-0.44437635	9.21099	0.6948662	0.07543882	0
TCF3	213730_x_at	Hs.371282.1.8	-0.08734955	8.46604	1.5840497	0.18710634	0
TIMM9	218316_at	Hs.440525.1.3	-0.44603154	7.7761397	0.6334871	0.0814655	0
C11orf24	52164_at	Hs.303025.1.2	-0.7289256	7.68625	0.7566761	0.098445415	0
BGN	201262_s_at	Hs.821.1.15	2.2126715	7.4992037	2.5747063	0.34333062	0
AP2A2	215764_x_at	Hs.19121.1.49	-0.79845464	7.378787	1.0894847	0.14765091	0
SLC16A3	217691_x_at	Hs.500761.1.20	-0.13807549	7.246716	4.043625	0.5579941	0
MGAT4A	219797_at	Hs.177576.1.19	0.031839963	7.2032847	0.5456244	0.0757466	0
ETAA1	219216_at	Hs.353022.1.14	0.015576937	6.9223785	1.0220212	0.14764017	0
MGAT2	203102_s_at	Hs.93338.1.10	-0.46199802	6.847179	1.5886164	0.23201035	0
TOP2A	201291_s_at	Hs.156346.1.29	-0.20762259	6.7122893	0.9236032	0.13759883	0
STMN1	200783_s_at	Hs.209983.1.5	-1.7571659	6.7105546	2.8431888	0.4236891	0
NAP1L1	204528_s_at	Hs.524599.1.9	-0.56638455	6.438269	0.88135374	0.13689296	0
SNX11	53912_at	Hs.15827.1.19	-0.1808786	6.425423	1.350209	0.21013542	0
ADRBK1	38447_at	Hs.83636.1.51	-0.73117715	6.405118	1.8479233	0.28850728	0
NAP1L1	204528_s_at	Hs.524599.1.11	-0.56417	6.3738985	0.6884583	0.108012125	0
LYPLA2	215568_x_at	Hs.533479.1.18	-1.8053951	6.3267455	1.0480416	0.16565256	0
DKC1	216212_s_at	Hs.4747.1.24	2.3276074	6.2315917	0.4562137	0.07320982	0
CDC6	203967_at	Hs.405958.1.26	-0.21053804	6.086678	0.46271598	0.076021105	0
PSMC6	201699_at	Hs.156171.1.27	-0.45245913	6.0183287	1.2953672	0.21523704	0
PSMC6	201699_at	Hs.156171.1.26	-0.4546106	6.0183287	1.2953672	0.21523704	0
AURKA	204092_s_at	Hs.250822.1.2	0.16784091	5.974195	0.79987586	0.13388847	0
TCF3	213730_x_at	Hs.371282.1.10	-0.08554339	5.9585114	0.8477694	0.14227872	0
HSPA9	200691_s_at	Hs.184233.1.8	0.5438669	5.92778	0.3615086	0.060985494	0
MMP7	204259_at	Hs.2256.1.3	-0.69102776	5.8967667	5.894563	0.99962634	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.8152903	5.720475	1.0901649	0.19057244	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.81769943	5.720475	1.0901649	0.19057244	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.82106376	5.720475	1.0901649	0.19057244	0
SEC22A	218703_at	Hs.477361.1.19	0.27951458	5.620965	0.5793693	0.10307293	0
SEC22A	218703_at	Hs.477361.1.21	0.28099138	5.599696	0.8174287	0.14597733	0
LRCH3	214739_at	Hs.518414.1.33	0.33499023	5.560019	0.45368433	0.08159762	0
ETAA1	219216_at	Hs.353022.1.16	0.016889581	5.536846	0.72445524	0.13084258	0
TIMM8A	205217_at	Hs.447877.1.3	1.555674	5.4936004	0.5089189	0.0926385	0
MMP7	204259_at	Hs.2256.1.7	-0.68887657	5.4235106	5.809583	1.071185	0
BCLAF1	214499_s_at	Hs.486542.1.16	0.6897759	5.358898	0.57514334	0.10732492	0
UBA5	218289_s_at	Hs.170737.1.21	0.29410404	5.342724	0.36212254	0.06777864	0
DCK	203302_at	Hs.709.1.27	0.36947405	5.22128	0.7451329	0.14271078	0
POLR2C	216282_x_at	Hs.79402.1.16	-0.28424302	5.2034473	0.7940595	0.15260258	0
C8orf39	51228_at	Hs.192788.1.2	1.0162911	4.9858336	0.893419	0.1791915	0
SLC46A3	214719_at	Hs.117167.1.1	-0.5197137	4.947672	0.94269097	0.19053224	0
GSTM4	210912_x_at	Hs.348387.1.24	-1.2762744	4.9042654	0.5166906	0.10535535	0
YME1L1	216304_x_at	Hs.499145.1.10	-0.8976959	4.895095	1.871449	0.38231108	0
CEP76	52285_f_at	Hs.236940.1.5	-0.12285543	4.888645	0.51774955	0.105908595	0
FAM21B	214946_x_at	Hs.365286.1.35	-0.87893504	4.876281	0.6828325	0.14003141	0
LPCAT1	201818_at	Hs.368853.1.3	0.44985366	4.871491	0.4366752	0.08963892	0
QPCT	205174_s_at	Hs.79033.1.18	0.003884405	4.795341	0.43948555	0.091648445	0
YARS2	218470_at	Hs.505231.1.6	-0.6149618	4.7719717	0.4732014	0.09916266	0
SET	213047_x_at	Hs.436687.1.19	1.2579172	4.763696	1.0172708	0.21354653	0
KTELC1	218587_s_at	Hs.231750.1.16	0.27268472	4.743526	0.25473773	0.05370219	0
YARS2	218470_at	Hs.505231.1.4	-0.6171523	4.733806	0.59276986	0.12522057	0
GINS3	45633_at	Hs.47125.1.11	-0.2699034	4.725053	0.53630173	0.11350174	0
CSTF1	32723_at	Hs.172865.1.25	0.17125405	4.66666	0.6389096	0.1369094	0
QKI	214543_x_at	Hs.510324.1.14	0.7177115	4.6602883	0.3258847	0.06992801	0
QKI	214543_x_at	Hs.510324.1.13	0.7154247	4.6602883	0.3258847	0.06992801	0
QKI	214543_x_at	Hs.510324.1.12	0.71261567	4.6602883	0.3258847	0.06992801	0
SNX3	213545_x_at	Hs.12102.1.3	0.6645959	4.6454782	0.6758001	0.14547482	0
FDX1	203647_s_at	Hs.744.1.6	-0.67439497	4.6240416	0.42418778	0.09173529	0
SIAH1	202981_x_at	Hs.295923.1.4	-0.289238	4.5784416	0.325765	0.07115194	0
POLE	216026_s_at	Hs.524871.1.3	-0.53121126	4.569564	0.56140006	0.12285638	0
ARF6	203312_x_at	Hs.525330.1.12	-0.458209	4.4482794	0.6362238	0.14302695	0
CCPG1	214151_s_at	Hs.285051.1.2	-0.37149575	4.4354715	0.3387587	0.0763749	0
SLC25A36	201919_at	Hs.144130.1.32	0.29883915	4.4018264	0.28819776	0.06547231	0

Table B.11 (continued)

Gene name	Probeset ID	PolyA Site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
CREBBP	202160_at	Hs.459759.1.9	-0.314064	4.37965	0.47235167	0.10785146	0
SLAMF8	219385_at	Hs.438683.1.13	-1.143719	4.2922716	0.21406662	0.049872573	0
SNX3	200067_x_at	Hs.12102.1.3	0.6621202	4.2461963	0.4237218	0.099788554	0
SERPINB8	206034_at	Hs.368077.1.9	-0.09324817	4.2455287	0.40142727	0.09455295	0
EXOSC8	215136_s_at	Hs.294041.1.21	-0.51023495	4.2224617	0.3999554	0.09472091	0
GRK6	202849_x_at	Hs.235116.1.40	0.5685016	4.1807356	0.7361597	0.17608377	0
COL16A1	204345_at	Hs.368921.1.2	-1.6539707	4.1697216	0.83637655	0.20058331	0
STK3	211078_s_at	Hs.492333.1.29	1.0350115	4.138397	0.8763039	0.21174958	0
TBC1D15	218268_at	Hs.284630.1.31	-0.5683697	4.1357336	0.3942492	0.09532751	0
CBX7	212914_at	Hs.356416.1.3	0.22028935	4.1101456	0.5619949	0.13673358	0
PLXND1	212235_at	Hs.301685.1.3	0.2893314	4.10435	0.3976072	0.09687459	0
CDC6	203968_s_at	Hs.405958.1.22	-0.20924345	4.103766	0.3398919	0.08282439	0
C12orf4	218374_s_at	Hs.302977.1.4	-0.64641094	4.0547295	0.18815696	0.04640432	0
GNS	203676_at	Hs.334534.1.8	-0.5772321	4.019027	0.20255876	0.050399948	0
ATP7A	205198_s_at	Hs.496414.1.21	1.5450938	3.9963508	0.32615054	0.08161209	0
VASP	202205_at	Hs.515469.1.28	-0.04082387	3.9730663	0.64723015	0.16290444	0
COL16A1	204345_at	Hs.368921.1.3	-1.6387897	3.8760114	0.5984247	0.15439188	0
KLHL20	210634_at	Hs.495035.1.12	-1.094665	3.8261528	0.18922067	0.04945455	0
VEGFA	210512_s_at	Hs.73793.1.28	0.6238738	3.7928944	0.80725646	0.2128339	0
C11orf41	214772_at	Hs.502266.1.7	-0.75713456	3.76237	0.2353257	0.06254719	0
BBS10	219487_at	Hs.96322.1.2	-0.56233567	3.7378037	0.32915425	0.08806087	0
ENOSF1	213645_at	Hs.369762.2.6	-0.12931003	3.7054071	0.3198744	0.08632638	0
HMGCS1	221750_at	Hs.397729.1.2	0.46858808	3.6939516	0.365757	0.0990151	0
ACTL6A	202666_s_at	Hs.435326.1.30	0.31946304	3.6801295	0.5703769	0.15498826	0
BBS10	219487_at	Hs.96322.1.3	-0.5605443	3.657374	0.40342712	0.11030514	0
CYP4A11	207407_x_at	Hs.1645.1.4	-1.4568208	3.6562562	0.25564682	0.06992038	0
LMO7	202674_s_at	Hs.207631.1.53	-0.49298128	3.6426153	1.2871115	0.3533482	0
CHAF1A	214426_x_at	Hs.79018.1.40	-0.076536074	3.6313186	0.388533	0.10699502	0
C10orf119	217905_at	Hs.124246.1.4	-0.83278394	3.6230679	0.58467364	0.1613753	0
CHGB	204260_at	Hs.516874.1.16	0.13062538	3.599309	0.18902636	0.052517403	0
PLSCR1	202430_s_at	Hs.130759.1.13	0.3037869	3.5976129	0.5942681	0.165184	0
NIPBL	212483_at	Hs.481927.1.71	0.46538222	3.5840802	0.43815112	0.12224925	0
METTL9	217868_s_at	Hs.279583.1.15	-0.29922277	3.5820265	1.2260888	0.34228915	0
AMOT	209521_s_at	Hs.528051.1.1	1.7136874	3.571619	1.387433	0.38846055	0
SCARB	2201647_s_at	Hs.349656.1.9	0.38143486	3.569423	0.2750423	0.07705511	0
DIRAS3	215506_s_at	Hs.194695.1.2	-1.3800749	3.5412014	0.23129594	0.06531567	0
TPI1	210050_at	Hs.524219.1.15	-0.6392035	3.4925218	1.2330604	0.35305732	0
RPL3	211666_x_at	Hs.119598.1.45	0.22179553	3.4778752	0.21210027	0.060985588	0
CSTF1	32723_at	Hs.172865.1.23	0.16975118	3.4363613	0.3510667	0.10216234	0
ITGB3BP	205176_s_at	Hs.166539.1.10	-1.4035974	3.4318025	0.22873628	0.06665193	0
40066	214720_x_at	Hs.469615.1.9	0.042242583	3.4270241	0.16752386	0.04888319	0
LETM1	218939_at	Hs.120165.1.4	0.3366399	3.4017751	0.33864045	0.09954816	0
USP9X	201100_s_at	Hs.77578.1.84	1.3963722	3.3909698	0.42767406	0.12612146	0

Table B.12 Significantly Lengthened Genes in Gastric Cardia Adenocarcinoma

Gene name	Probeset ID	PolyA site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
RFC3	204127_at	Hs.115474.1.25	-0.5141241	-7.980033	-0.97700167	0.12243078	0
CEP164	204250_s_at	Hs.504009.1.31	-0.662535	-7.963188	-1.1031882	0.13853599	0
GPR107	211977_at	Hs.512461.1.34	1.2706476	-7.730407	-0.6455587	0.08350902	0
PHTF2	215286_s_at	Hs.203965.1.19	0.7853471	-7.5035176	-1.2780008	0.17032023	0
PHTF2	215286_s_at	Hs.203965.1.18	0.7827358	-7.5035176	-1.2780008	0.17032023	0
INTS8	218905_at	Hs.521693.1.33	1.0233246	-6.7171216	-1.6482725	0.24538375	0
PPIG	208994_s_at	Hs.470544.1.19	0.0693969	-6.679842	-0.62932086	0.094211936	0
AQP4	210906_x_at	Hs.315369.1.10	-0.116910934	-6.6033807	-0.6503562	0.09848837	0
SF3A2	209381_x_at	Hs.115232.1.23	-0.081058964	-5.9929132	-0.95438015	0.15925145	0
FABP4	203980_at	Hs.391561.1.9	0.9948256	-5.8283277	-0.4614917	0.07918081	0
ISL1	206104_at	Hs.505.1.14	0.47260508	-5.7132993	-0.6250998	0.10941135	0
TTLL12	216251_s_at	Hs.517670.1.4	0.23271485	-5.5670156	-0.4059639	0.07292307	0
ACTR8	218658_s_at	Hs.412186.1.3	0.25816894	-5.426888	-0.3349154	0.06171408	0
RAB40B	204547_at	Hs.484068.1.2	-0.13669431	-5.313349	-0.52779317	0.099333435	0
JRKL	206734_at	Hs.105940.1.2	-0.6937938	-5.198997	-0.8709693	0.16752641	0
C9orf78	218116_at	Hs.278429.1.3	1.2640687	-5.1518326	-0.39861286	0.07737302	0
PRSS21	220051_at	Hs.72026.1.3	-0.3173186	-4.9815216	-0.74585986	0.1497253	0
SFRS6	206108_s_at	Hs.6891.1.21	0.1566716	-4.9609275	-0.48643076	0.09805238	0
C13orf18	219471_at	Hs.98117.1.4	-0.5022415	-4.879705	-0.4055593	0.08311144	0
CLU	222043_at	Hs.436657.1.7	0.93649375	-4.868297	-6.0048904	1.2334684	0
FAM118A	219629_at	Hs.265018.1.19	0.23414396	-4.851004	-0.30887568	0.06367253	0
PVR	32699_s_at	Hs.171844.1.18	-0.044020697	-4.8198347	-0.5768173	0.11967574	0
LARP1	212193_s_at	Hs.292078.1.54	0.5619869	-4.7588625	-0.364246	0.07654056	0
C19orf42	219097_x_at	Hs.356467.1.21	-0.06228147	-4.730584	-0.55973446	0.118322484	0
RAD23B	201223_s_at	Hs.521640.1.23	1.2015281	-4.729788	-0.9094062	0.19227208	0
RUSC1	206949_s_at	Hs.226499.1.30	-1.1639885	-4.6746	-0.33982253	0.07269553	0
UNKNOWN	201254_x_at	Hs.408073.1.41	1.0787921	-4.6356506	-0.22746444	0.049068503	0
GSTM5	205752_s_at	Hs.75652.1.12	-1.2515329	-4.5686727	-0.34215903	0.07489244	0
TREM1	219434_at	Hs.283022.1.8	0.61678636	-4.5282497	-0.7779745	0.17180468	0
GLUL	215001_s_at	Hs.518525.1.4	-1.0395415	-4.477539	-0.40046155	0.08943787	0
CD44	210916_s_at	Hs.502328.1.11	-0.7499632	-4.475213	-0.3612795	0.080729	0
CD44	210916_s_at	Hs.502328.1.28	-0.74501324	-4.475213	-0.3612795	0.080729	0
CD44	210916_s_at	Hs.502328.1.13	-0.747614	-4.475213	-0.3612795	0.080729	0
CHD4	201184_s_at	Hs.162233.1.4	-0.6416917	-4.4669795	-0.4611975	0.10324594	0
SNRPD2	200826_at	Hs.515472.1.3	-0.039374575	-4.4531283	-0.40691745	0.091377884	0
NLK	218318_s_at	Hs.208759.1.27	-0.22178558	-4.440776	-0.22774339	0.051284593	0
ZKSCAN5	203731_s_at	Hs.110839.1.21	0.8210271	-4.4073906	-0.18246639	0.041400094	0
PSAP	200871_s_at	Hs.523004.1.3	-0.86388934	-4.3729944	-0.585263	0.13383576	0
PLCG1	202789_at	Hs.268177.1.75	0.15375715	-4.344072	-0.99346566	0.22869457	0
ADI1	217761_at	Hs.502773.1.3	-0.023799473	-4.2984123	-0.3346219	0.07784779	0
BICD1	214806_at	Hs.505202.1.23	-0.6193577	-4.214725	-0.36178517	0.085838385	0
ACTR8	218658_s_at	Hs.412186.1.4	0.25959536	-4.214325	-0.43442202	0.10308223	0
UCHL1	201387_s_at	Hs.518731.1.17	0.35790837	-4.161666	-0.24937809	0.059922658	0
FGB	204988_at	Hs.300774.1.15	0.42787325	-4.1414104	-0.4682896	0.11307491	0
RRM2	201890_at	Hs.226390.1.19	-0.020591544	-4.123335	-0.20573139	0.04989442	0
ANXA6	200982_s_at	Hs.412117.1.7	0.5582834	-4.116355	-0.32614803	0.079232246	0
WDR12	218512_at	Hs.73291.1.3	0.09665224	-4.092773	-0.22780073	0.055659264	0
PILRA	222218_s_at	Hs.444407.1.4	0.8294333	-4.0760436	-0.31906593	0.07827834	0
LYPLA1	203007_x_at	Hs.435850.1.12	0.96993023	-4.0589023	-0.35621488	0.08776139	0
KYNU	210663_s_at	Hs.470126.1.18	0.06085765	-4.0498004	-0.48560572	0.11990856	0
KYNU	210663_s_at	Hs.470126.1.16	0.057610087	-4.0498004	-0.48560572	0.11990856	0
KYNU	210663_s_at	Hs.470126.1.17	0.05924461	-4.0498004	-0.48560572	0.11990856	0
AP3D1	208710_s_at	Hs.512815.1.32	-0.08243066	-4.020296	-0.25593436	0.06366058	0
UBXN7	212840_at	Hs.518524.1.3	0.33178157	-4.011551	-0.26241028	0.065413676	0
RBBP8	203344_s_at	Hs.546282.1.47	-0.118532546	-4.005135	-0.3534689	0.08825392	0
FBXO17	220233_at	Hs.531770.1.10	-0.04871134	-3.994175	-0.48673272	0.12186064	0
SPP1	209875_s_at	Hs.313.1.13	0.39542207	-3.9909408	-0.27526653	0.06897284	0
WAS	38964_r_at	Hs.2157.1.17	1.4215444	-3.973154	-0.6262362	0.1576169	0
FMO6P	215174_at	Hs.448988.1.3	-1.0990516	-3.9382277	-0.46838975	0.11893415	0
CRYZ	202950_at	Hs.83114.1.5	-1.3472788	-3.8932679	-0.39371717	0.101127684	0
EIF5B	201024_x_at	Hs.158688.1.13	0.034774706	-3.8741004	-0.43918133	0.11336343	0
TRAPPC10	215269_at	Hs.126221.1.16	0.19894141	-3.8578174	-0.36698675	0.095128074	0

Table B.12 (continued)

Gene name	Probeset ID	PolyA site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
GOSR2	213144_at	Hs.463278.1.29	-0.1902358	-3.8473318	-0.46201444	0.120086975	0
TTK	204822_at	Hs.169840.1.41	0.6414308	-3.8279195	-0.16029525	0.04187529	0
CCL13	206407_s_at	Hs.414629.1.3	-0.21388143	-3.8258789	-0.30787253	0.08047106	0
UNKNOWN	201254_x_at	Hs.408073.1.40	1.0752367	-3.815044	-0.18621802	0.048811503	0
MRPL3	208787_at	Hs.205163.1.4	0.29243654	-3.807298	-0.33574688	0.08818508	0
PCYT1A	204210_s_at	Hs.435767.1.19	0.3281985	-3.80567	-0.34483743	0.09061149	0
PCYT1A	204210_s_at	Hs.435767.1.20	0.33001217	-3.80567	-0.34483743	0.09061149	0
TMEM14A	218477_at	Hs.94896.1.7	0.63440573	-3.7868063	-0.44646418	0.117899925	0
PPIC	204517_at	Hs.110364.1.4	0.52588916	-3.769874	-0.4315611	0.11447626	0
CABC1	218168_s_at	Hs.118241.1.35	-0.9759909	-3.7101448	-0.39283514	0.10588135	0
GSTM1	215333_x_at	Hs.301961.1.12	-1.2644639	-3.700206	-0.18150628	0.049053017	0
HDGFRP3	209524_at	Hs.513954.1.4	-0.33090046	-3.6294134	-0.5072149	0.13975121	0
PTMA	200772_x_at	Hs.459927.1.1	0.116673164	-3.6263587	-0.22972167	0.06334775	0
HSPA4L	205543_at	Hs.135554.1.18	0.4172134	-3.6142862	-0.5588095	0.1546113	0
COX4I1	202698_x_at	Hs.433419.1.14	-0.24874565	-3.607966	-0.1721636	0.047717635	0
NFIC	213298_at	Hs.170131.1.21	-0.07964809	-3.5911179	-0.28330457	0.07889036	0
FBXO28	202272_s_at	Hs.64691.1.16	-0.98360986	-3.5468953	-0.20033467	0.056481697	0
HMOX2	218121_at	Hs.284279.1.14	-0.31030205	-3.5438273	-0.3850696	0.108659245	0
TPD52	201690_s_at	Hs.368433.1.8	0.9909694	-3.5377722	-0.5428643	0.15344808	0
SHOX	2208443_x_at	Hs.55967.1.5	0.3086395	-3.5262055	-0.3163092	0.089702435	0
CD84	205988_at	Hs.398093.1.2	-1.1238027	-3.5168364	-0.23416209	0.066583164	0
SNX13	213292_s_at	Hs.487648.1.5	0.7491782	-3.4887488	-0.391338	0.11217144	0
CSK	202329_at	Hs.77793.1.24	-0.34222883	-3.4848118	-0.20302558	0.058260128	0

Table B.13 Significantly Shortened Genes in Gastric Noncardia Adenocarcinoma

Gene name	Probeset ID	PolyA site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
BGN	201262_s_at	Hs.821.1.15	2.2462237	8.902401	3.6349754	0.40831405	0
TIMM9	218316_at	Hs.440525.1.3	-0.45028946	8.522291	0.72103715	0.08460602	0
GLRX3	214205_x_at	Hs.42644.1.17	-0.82553476	8.1270075	1.496742	0.1841689	0
GLRX3	214205_x_at	Hs.42644.1.16	-0.82806265	8.1270075	1.496742	0.1841689	0
GLRX3	214205_x_at	Hs.42644.1.14	-0.831001	8.1270075	1.496742	0.1841689	0
SNX11	53912_at	Hs.15827.1.19	-0.18420371	7.4706144	1.8087366	0.2421135	0
AP2A2	215764_x_at	Hs.19121.1.49	-0.8087534	7.4227343	0.86903715	0.11707777	0
SEC22A	218703_at	Hs.477361.1.21	0.286971	7.1288586	0.9101784	0.12767519	0
TIMM9	218316_at	Hs.440525.1.4	-0.44858518	7.0267673	0.5542135	0.078871764	0
NAP1L1	204528_s_at	Hs.524599.1.9	-0.57640314	6.7466207	1.2474957	0.18490674	0
STMN1	200783_s_at	Hs.209983.1.5	-1.761871	6.6276317	2.5774355	0.38889238	0
CDC6	203967_at	Hs.405958.1.26	-0.21349367	6.456685	0.63127637	0.097770974	0
MGAT4A	219797_at	Hs.177576.1.19	0.03883447	6.4135303	0.5060879	0.07890941	0
ETAA1	219216_at	Hs.353022.1.14	0.020000136	6.3847766	1.1302993	0.17703037	0
ATP7A	205198_s_at	Hs.496414.1.21	1.5491467	6.31469	0.46286607	0.073299885	0
POLR2C	216282_x_at	Hs.79402.1.16	-0.28305522	6.2567577	0.8484837	0.13561076	0
CCPG1	214151_s_at	Hs.285051.1.2	-0.3751269	6.1946845	0.39629388	0.06397321	0
TOP2A	201291_s_at	Hs.156346.1.29	-0.21034037	6.0360565	0.73199916	0.12127109	0
SNX3	213545_x_at	Hs.12102.1.3	0.6684582	5.9719944	0.8355751	0.13991559	0
TDRKH	221052_at	Hs.144439.1.4	-1.1952298	5.909816	0.305362	0.051670305	0
GNS	203676_at	Hs.334534.1.8	-0.5863981	5.7660666	0.33171403	0.057528652	0
PSMC6	201699_at	Hs.156171.1.27	-0.4574974	5.757143	1.1481352	0.19942795	0
PSMC6	201699_at	Hs.156171.1.26	-0.45961782	5.757143	1.1481352	0.19942795	0
COL16A1	204345_at	Hs.368921.1.2	-1.6323656	5.749716	1.094193	0.19030383	0
MMP7	204259_at	Hs.2256.1.3	-0.6952467	5.725717	3.4272382	0.5985693	0
COL16A1	204345_at	Hs.368921.1.3	-1.6192032	5.7194085	0.8698969	0.1520956	0
CEP76	52285_f_at	Hs.236940.1.5	-0.122597165	5.685023	0.73665833	0.12957878	0
BCLAF1	214499_s_at	Hs.486542.1.16	0.6899825	5.6417437	0.61828136	0.10959047	0
MMP7	204259_at	Hs.2256.1.7	-0.6924209	5.635337	3.22687	0.57261354	0
POLE	216026_s_at	Hs.524871.1.3	-0.5392182	5.6220093	0.54816246	0.09750295	0
RBMS	234187_at	Hs.505729.1.16	-0.5966936	5.601369	0.38975143	0.069581464	0
LYPLA2	215568_x_at	Hs.533479.1.18	-1.8095484	5.569804	1.0330055	0.1854653	0
SNX3	200067_x_at	Hs.12102.1.3	0.6661844	5.38021	0.5536759	0.10290972	0
ADRBK1	38447_at	Hs.83636.1.51	-0.7346622	5.3190513	1.1604276	0.21816438	0
MGAT2	203102_s_at	Hs.93338.1.10	-0.46706343	5.3098793	1.251008	0.23560008	0
DKC1	216212_s_at	Hs.4747.1.24	2.3660498	5.2629957	0.42186832	0.08015745	0
RABGEF1	218310_at	Hs.530053.1.33	0.78249866	5.261423	1.7785242	0.338031	0
TIMM8A	205217_at	Hs.447877.1.3	1.5623424	5.070706	0.5566691	0.109781384	0
SLC16A3	217691_x_at	Hs.500761.1.20	-0.1385046	5.0632734	2.3939102	0.4727989	0
ATP6V1C1	202872_at	Hs.86905.1.21	1.028364	5.0576725	0.44037294	0.08707028	0
COL15A1	203477_at	Hs.409034.1.48	1.194968	5.0422907	0.36039007	0.07147348	0
COL15A1	203477_at	Hs.409034.1.47	1.1892186	5.0422907	0.36039007	0.07147348	0
UNKNOWN	216304_x_at	Hs.499145.1.10	-0.9094913	5.014855	3.2484188	0.64775926	0
VASP	202205_at	Hs.515469.1.28	-0.03939237	4.9812455	0.9231694	0.18532902	0
ADSL	210250_x_at	Hs.75527.1.19	0.23479193	4.9637184	0.7386081	0.14880137	0
C11orf24	52164_at	Hs.303025.1.2	-0.7318491	4.9503317	0.5265385	0.10636429	0
ARF6	203312_x_at	Hs.525330.1.12	-0.46318138	4.913985	0.47789574	0.09725218	0
LPCAT1	201818_at	Hs.368853.1.3	0.45425272	4.8936963	0.6213962	0.1269789	0
AURKA	204092_s_at	Hs.250822.1.2	0.17754963	4.8894944	0.4667833	0.09546658	0
FDX1	203647_s_at	Hs.744.1.6	-0.67972064	4.861984	0.33787644	0.06949353	0
DCK	203302_at	Hs.709.1.27	0.37914842	4.8291903	0.7008834	0.14513476	0
HSPA9	200691_s_at	Hs.184233.1.8	0.54647774	4.822292	0.3801663	0.07883519	0
SF4	209547_s_at	Hs.515274.1.3	-0.05453651	4.7881784	0.49599743	0.10358792	0
CYP4A11	207407_x_at	Hs.1645.1.4	-1.4499975	4.786864	0.3363787	0.07027121	0
QPCT	205174_s_at	Hs.79033.1.18	0.00739058	4.776507	0.73176014	0.15319985	0
TIMM8A	205217_at	Hs.447877.1.4	1.5725039	4.7232356	0.39817286	0.08430087	0
RPL3	211666_x_at	Hs.119598.1.45	0.23156385	4.7052155	0.21867406	0.046474826	0
BBS10	219487_at	Hs.96322.1.2	-0.5725621	4.573056	0.57123506	0.12491319	0
ETAA1	219216_at	Hs.353022.1.16	0.021409988	4.54161	0.79780364	0.1756654	0
GPR183	205419_at	Hs.784.1.2	-0.49802983	4.4992995	0.43761706	0.09726337	0
LDB1	35160_at	Hs.454418.1.3	-0.8535017	4.4897103	1.247664	0.27789408	0

Table B.13 (continued)

Gene name	Probeset ID	PolyA site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
CREBBP	202160_at	Hs.459759.1.9	-0.3185633	4.4724727	0.49292886	0.11021394	0
C9orf156	47530_at	Hs.9196.1.2	1.1646284	4.4351296	0.8876518	0.20014112	0
PLXND1	212235_at	Hs.301685.1.3	0.2971995	4.431139	0.47351992	0.1068619	0
SEC22A	218703_at	Hs.477361.1.19	0.28504205	4.4114137	0.44896078	0.10177254	0
ESF1	218859_s_at	Hs.369284.1.19	0.14336304	4.400385	0.33968997	0.07719551	0
KTELC1	218587_s_at	Hs.231750.1.16	0.2780549	4.395982	0.22825634	0.05192386	0
LRCH3	214739_at	Hs.518414.1.33	0.3407701	4.3794346	0.2854705	0.065184325	0
HSD17B7	220081_x_at	Hs.492925.1.21	-1.1097809	4.372202	0.3502401	0.08010612	0
TCF3	213730_x_at	Hs.371282.1.8	-0.08733765	4.3350253	0.8506017	0.19621608	0
WDHD1	204727_at	Hs.385998.1.1	-0.4540964	4.326763	0.25088286	0.05798396	0
FAM21B	214946_x_at	Hs.365286.1.35	-0.8928776	4.320964	0.63032246	0.14587542	0
EXOSC8	215136_s_at	Hs.294041.1.21	-0.5173931	4.293646	0.37241924	0.0867373	0
GINS3	45633_at	Hs.47125.1.11	-0.2682951	4.145379	0.6032697	0.14552823	0
C1QB	202953_at	Hs.8986.1.7	-1.8673699	4.132119	0.7172117	0.17356995	0
HNRNPA1	214280_x_at	Hs.546261.1.20	-0.60466826	4.110387	0.32358682	0.078724176	0
HNRNPA1	214280_x_at	Hs.546261.1.23	-0.6027326	4.110387	0.32358682	0.078724176	0
HNRNPA1	214280_x_at	Hs.546261.1.24	-0.600969	4.110387	0.32358682	0.078724176	0
BBS10	219487_at	Hs.96322.1.3	-0.5705105	4.0873294	0.7159152	0.17515476	0
YARS2	218470_at	Hs.505231.1.6	-0.6185687	4.0659604	0.5056584	0.124363825	0
MLPH	218211_s_at	Hs.102406.1.41	0.12905498	4.0412936	0.41677415	0.103128895	0
NAP1L1	204528_s_at	Hs.524599.1.11	-0.5740245	4.0353656	0.26735127	0.06625206	0
SIAH1	202981_x_at	Hs.295923.1.4	-0.2880951	4.0264297	0.27048266	0.0671768	0
ZAK	218833_at	Hs.444451.1.34	0.08240429	3.9843075	0.53346145	0.13389063	0
SERPINB8	206034_at	Hs.368077.1.9	-0.09476923	3.914074	0.4566269	0.116662815	0
C10orf119	217905_at	Hs.124246.1.4	-0.84190714	3.8922515	0.7141018	0.18346754	0
KCNN1	206231_at	Hs.158173.1.9	-0.057384487	3.8880017	0.40384626	0.10386988	0
SIK1	208078_s_at	Hs.282113.1.2	0.20243685	3.8705904	0.2305088	0.059553914	0
GSTM4	210912_x_at	Hs.348387.1.24	-1.2656091	3.8195143	0.44099963	0.115459606	0
SLC25A36	201919_at	Hs.144130.1.32	0.30656457	3.8189936	0.22055769	0.05775283	0

Table B.14 Significantly Lengthened Genes in Gastric Noncardia Adenocarcinoma

Gene name	Probeset ID	PolyA site ID	Expected Score	Observed Score	Numerator (r)	Denominator (s)	q-value (%)
CEP164	204250_s_at	Hs.504009.1.31	-0.6667729	-10.612372	-1.1516268	0.10851738	0
IVL	214599_at	Hs.516439.1.3	-1.1851178	-9.353232	-1.0938075	0.11694433	0
PHTF2	215286_s_at	Hs.203965.1.18	0.7854199	-9.187118	-1.6523111	0.17985088	0
PHTF2	215286_s_at	Hs.203965.1.19	0.78764766	-9.187118	-1.6523111	0.17985088	0
FBXO17	220233_at	Hs.531770.1.10	-0.04707854	-9.01202	-1.030907	0.114392444	0
TREM1	219434_at	Hs.283022.1.8	0.62282866	-8.952754	-1.6763357	0.18724246	0
JRKL	206734_at	Hs.105940.1.2	-0.6980922	-8.8964	-1.5089116	0.16960923	0
AQP4	210906_x_at	Hs.315369.1.10	-0.116144836	-8.427729	-1.0530901	0.124955386	0
SF3A2	209381_x_at	Hs.115232.1.23	-0.081306696	-8.307971	-1.3211263	0.15901913	0
RAD23B	201223_s_at	Hs.521640.1.23	1.2091358	-8.057215	-2.579424	0.3201384	0
TFDP1	212330_at	Hs.79353.1.30	-0.48274097	-7.514747	-0.90250874	0.12009835	0
SNRPD2	200826_at	Hs.515472.1.3	-0.03747829	-6.805327	-0.7282747	0.107015386	0
TTLL12	216251_s_at	Hs.517670.1.4	0.24268535	-6.7408547	-0.7707479	0.11433979	0
CAND1	208838_at	Hs.546407.1.58	-0.5824982	-6.7142816	-0.5163133	0.07689778	0
GPR107	211977_at	Hs.512461.1.34	1.2682686	-6.6182733	-0.4281789	0.06469647	0
C13orf18	219471_at	Hs.98117.1.4	-0.50971764	-6.52103	-0.47276402	0.072498366	0
CDC42EP3	209288_s_at	Hs.369574.1.3	0.009098717	-6.4974313	-0.47523475	0.07314194	0
TPD52	201690_s_at	Hs.368433.1.8	0.9806003	-6.3593106	-0.9141729	0.14375345	0
RAB40B	204547_at	Hs.484068.1.2	-0.137088	-6.265147	-0.5104021	0.08146689	0
NFIC	213298_at	Hs.170131.1.21	-0.07975614	-6.2481704	-0.421381	0.0674407	0
CHD4	201184_s_at	Hs.162233.1.4	-0.64622086	-6.2472463	-0.78302383	0.12533903	0
C19orf42	219097_x_at	Hs.356467.1.21	-0.06149355	-6.120473	-0.6110921	0.099843934	0
TRAPPC10	215269_at	Hs.126221.1.16	0.20891269	-5.991048	-0.5592973	0.09335551	0
HMOX2	218121_at	Hs.284279.1.14	-0.3155273	-5.9227853	-0.4724412	0.07976673	0
FMO6P	215174_at	Hs.448988.1.3	-1.0969007	-5.874056	-0.91592574	0.15592732	0
PPIG	208994_s_at	Hs.470544.1.19	0.077922285	-5.870356	-0.80240655	0.13668789	0
LARP1	212193_s_at	Hs.292078.1.54	0.5658026	-5.806447	-0.368459	0.06345688	0
INTS8	218905_at	Hs.521693.1.33	1.0088601	-5.801202	-1.5174146	0.261569	0
FAM118A	219629_at	Hs.265018.1.19	0.24449322	-5.7665215	-0.33123565	0.057441153	0
PVR	32699_s_at	Hs.171844.1.18	-0.04256729	-5.708096	-0.8373549	0.146696	0
RFC3	204127_at	Hs.115474.1.25	-0.52108866	-5.682369	-0.84313726	0.14837776	0
AKT3	212607_at	Hs.498292.1.7	-0.9619059	-5.48667	-0.37781715	0.06886092	0
RRM2	201890_at	Hs.226390.1.19	-0.020827504	-5.3939986	-0.32749355	0.06071443	0
METTL9	217868_s_at	Hs.279583.1.14	-0.30122578	-5.2740374	-0.6854452	0.12996593	0
PTMA	200772_x_at	Hs.459927.1.1	0.12325163	-5.2693925	-0.3628075	0.068851866	0
RAD21	200608_s_at	Hs.81848.1.3	1.0424874	-5.2532563	-0.3748417	0.07135416	0
PILRA	222218_s_at	Hs.444407.1.4	0.8330783	-5.2511945	-0.3929546	0.07483147	0
PPIC	204517_at	Hs.110364.1.4	0.52914345	-5.2502723	-0.5720676	0.10895961	0
FABP4	203980_at	Hs.391561.1.9	0.98376095	-5.217928	-0.33843756	0.06486053	0
BICD1	214806_at	Hs.505202.1.23	-0.62276655	-5.1828523	-0.40471613	0.07808753	0
WIZ	52005_at	Hs.442138.1.3	-0.06292481	-5.1408777	-0.35776854	0.06959289	0
NNT	202783_at	Hs.482043.1.62	0.47575447	-5.0901723	-1.6303959	0.3203027	0
SFRS6	206108_s_at	Hs.6891.1.21	0.16660622	-5.0138884	-0.51947606	0.10360742	0
CLU	222043_at	Hs.436657.1.7	0.9296237	-5.0083685	-5.7396154	1.146005	0
FAM189B	203550_s_at	Hs.348308.1.7	-1.1604651	-4.9925976	-0.18945312	0.037946805	0
FGB	204988_at	Hs.300774.1.15	0.43294984	-4.9287014	-0.48285174	0.097967334	0
LAMP2	203042_at	Hs.496684.1.5	1.7597665	-4.8856072	-0.71923804	0.1472157	0
OSBPL2	209222_s_at	Hs.473254.1.38	0.18572803	-4.865187	-0.43418217	0.089242645	0
GLUL	215001_s_at	Hs.518525.1.4	-1.0475004	-4.817653	-0.32263505	0.06696934	0
40066	214720_x_at	Hs.469615.1.10	0.052114263	-4.7681074	-0.2956673	0.06200936	0
FZD10	219764_at	Hs.31664.1.1	-0.54137856	-4.758103	-1.2096467	0.25422877	0
FADS1	208962_s_at	Hs.503546.1.8	-0.7397526	-4.644609	-0.2401079	0.05169604	0
USP14	201672_s_at	Hs.464416.1.27	-0.13224223	-4.643683	-0.26379693	0.056807697	0
ACTR8	218658_s_at	Hs.412186.1.3	0.26475003	-4.6415124	-0.2907052	0.06263157	0
GNL3	217850_at	Hs.313544.1.20	0.2613361	-4.583144	-0.45725632	0.09976913	0
TACC1	217437_s_at	Hs.279245.1.41	0.95228964	-4.504904	-1.0457006	0.23212494	0
ZKSCAN5	203731_s_at	Hs.110839.1.21	0.8253085	-4.4508176	-0.18334854	0.04119435	0
OGFOD1	221090_s_at	Hs.231883.1.33	-0.28484297	-4.4278455	-0.38816178	0.08766381	0
EPS15	217886_at	Hs.83722.1.3	-1.432246	-4.4193025	-0.61361134	0.138848	0
PBX2	202875_s_at	Hs.509545.1.5	0.6079392	-4.4122295	-0.25222027	0.057163905	0
CCDC93	219774_at	Hs.107845.1.3	0.05704335	-4.3825045	-0.21806347	0.04975773	0

Table B.14 (continued)

Gene name	Probeset ID	PolyA site ID	Expected score	Observed score	Numerator (r)	Denominator (s)	q-value (%)
SKP1	200711_s_at	Hs.171626.1.7	0.5384926	-4.3619814	-0.48748374	0.11175741	0
DENND3	212974_at	Hs.18166.1.45	1.0558566	-4.3472505	-0.3933599	0.09048475	0
UBXN7	212840_at	Hs.518524.1.3	0.337507	-4.346662	-0.2419529	0.05566407	0
PCYT1A	204210_s_at	Hs.435767.1.20	0.33595875	-4.340528	-0.3537128	0.08149073	0
PCYT1A	204210_s_at	Hs.435767.1.19	0.33410877	-4.340528	-0.3537128	0.08149073	0
RPL31	221593_s_at	Hs.469473.1.5	0.047656357	-4.2913513	-0.67939186	0.15831654	0
C13orf18	219471_at	Hs.98117.1.7	-0.5076283	-4.257125	-0.40725863	0.09566518	0
SPP1	209875_s_at	Hs.313.1.13	0.3995383	-4.2460146	-0.3099146	0.07298952	0
C16orf35	210672_s_at	Hs.19699.1.10	-0.3238103	-4.2244463	-0.49366522	0.11685915	0
IK	200066_at	Hs.421245.1.28	0.5485396	-4.2193985	-0.3199618	0.07583114	0
TBC1D9	212956_at	Hs.480819.1.3	0.42427725	-4.198646	-0.2561648	0.061011285	0
DPP4	203717_at	Hs.368912.1.2	0.073306456	-4.180487	-0.28524053	0.06823141	0
FEZ2	215000_s_at	Hs.258563.1.6	0.006025521	-4.0773005	-0.21569109	0.052900463	0
CYP3A7	211843_x_at	Hs.111944.1.4	0.82805234	-4.056866	-2.000499	0.49311438	0
CYP3A7	211843_x_at	Hs.111944.1.5	0.8307421	-4.056866	-2.000499	0.49311438	0
ASB8	218841_at	Hs.432699.1.3	-0.61389554	-4.0354276	-0.405918	0.1005886	0
EHD3	218935_at	Hs.368808.1.13	0.004379981	-4.010436	-0.3029852	0.075549185	0
SFRS11	200686_s_at	Hs.479693.1.43	-1.3536211	-4.0035663	-1.812747	0.45278308	0
XYLT1	213725_x_at	Hs.22907.1.3	-0.3085581	-4.0022388	-0.5732703	0.14323741	0
NEIL3	219502_at	Hs.405467.1.23	0.44827995	-3.9878988	-0.5416808	0.13583113	0
TNC	201645_at	Hs.143250.1.6	1.2305567	-3.9773185	-0.23935771	0.060180675	0
MYO1B	212365_at	Hs.439620.1.46	0.095137656	-3.967899	-0.24132812	0.060820125	0
DLX5	213707_s_at	Hs.99348.1.4	0.8224723	-3.9621441	-0.4406581	0.111217074	0
LPL	203549_s_at	Hs.180878.1.36	0.91653025	-3.9357884	-0.66939497	0.17007901	0
ANXA6	200982_s_at	Hs.412117.1.7	0.56071186	-3.9306288	-0.27860868	0.070881456	0
RPL6	200034_s_at	Hs.528668.1.6	-0.55129296	-3.9171307	-0.17340004	0.044267107	0
FOXA	240284_at	Hs.155651.1.1	0.14962523	-3.9135752	-0.7617643	0.19464664	0
SRP72	208802_at	Hs.237825.1.39	0.37748417	-3.8632715	-0.99640465	0.2579173	0
PSAP	200871_s_at	Hs.523004.1.3	-0.87194854	-3.8620317	-0.46275496	0.11982164	0
CCL13	206407_s_at	Hs.414629.1.3	-0.21648541	-3.8263836	-0.36799037	0.09617185	0
GNS	203676_at	Hs.334534.1.5	-0.58825964	-3.8146262	-0.28302014	0.07419341	0
AP3D1	208710_s_at	Hs.512815.1.32	-0.082902804	-3.7919755	-0.18809962	0.04960465	0
CABC1	218168_s_at	Hs.118241.1.35	-0.97594666	-3.785677	-0.4361626	0.11521389	0
RAD17	207405_s_at	Hs.16184.1.34	0.49105418	-3.7775986	-0.17844713	0.047238246	0
UCHL1	201387_s_at	Hs.518731.1.17	0.36497977	-3.7746484	-0.27088475	0.07176423	0
B4GALT1	216627_s_at	Hs.272011.1.15	1.0840105	-3.7687273	-0.23861814	0.06331531	0
TMEM14A	218477_at	Hs.94896.1.7	0.6404868	-3.7617104	-0.3116436	0.08284625	0
SKP1	200711_s_at	Hs.171626.1.6	0.53692484	-3.7564983	-0.294425	0.078377515	0
SCG5	203889_at	Hs.156540.1.15	-0.39616856	-3.7481418	-0.24745452	0.066020586	0
CORO2A	205538_at	Hs.113094.1.7	1.1742692	-3.7197442	-0.37381792	0.1004956	0
GSTM1	215333_x_at	Hs.301961.1.12	-1.2522905	-3.7133787	-0.18702638	0.050365556	0
ARSJ	219973_at	Hs.22895.1.3	0.4172788	-3.7093654	-0.17899013	0.04825357	0
EIF5B	201024_x_at	Hs.158688.1.13	0.04202039	-3.708358	-0.61301756	0.165307	0
THBS2	203083_at	Hs.371147.1.4	0.7310388	-3.7048795	-0.26876962	0.07254477	0

Table B.15 Significantly Altered Genes in both Cardia and Noncardia Tumors

3'UTR Shortening			3' UTR Lengthening		
ATP6V1C1	*TIMM9*	*TCF3*	*RFC3*	*CEP164*	*GPR107*
C11orf24	*BGN*	*AP2A2*	*PHTF2*	*INTS8*	*PPIG*
SLC16A3	*MGAT4A*	*ETAA1*	*AQP4*	*SF3A2*	*FABP4*
MGAT2	*TOP2A*	*STMN1*	*TTLL12*	*ACTR8*	*RAB40B*
NAP1L1	*SNX11*	*ADRBK1*	*JRKL*	*SFRS6*	*C13orf18*
LYPLA2	*DKC1*	*CDC6*	*CLU*	*FAM118A*	*PVR*
PSMC6	*AURKA*	*HSPA9*	*LARP1*	*C19orf42*	*RAD23B*
MMP7	*GLRX3*	*SEC22A*	*TREM1*	*GLUL*	*CHD4*
LRCH3	*TIMM8A*	*BCLAF1*	*SNRPD2*	*ZKSCAN5*	*PSAP*
DCK	*POLR2C*	*GSTM4*	*BICD1*	*UCHL1*	*FGB*
YME1L1	*CEP76*	*FAM21B*	*RRM2*	*ANXA6*	*PILRA*
LPCAT1	*QPCT*	*YARS2*	*AP3D1*	*UBXN7*	*FBXO17*
KTELC1	*GINS3*	*SNX3*	*SPP1*	*FMO6P*	*EIF5B*
FDX1	*SIAH1*	*POLE*	*TRAPPC10*	*CCL13*	*PCYT1A*
ARF6	*CCPG1*	*SLC25A36*	*TMEM14A*	*PPIC*	*CABC1*
CREBBP	*SNX3* (different probeset)	*SERPINB8*	*GSTM1*	*PTMA*	*NFIC*
EXOSC8	*COL16A1*	*PLXND1*	*HMOX2*	*TPD52*	
GNS	*ATP7A*	*VASP*			
BBS10	*CYP4A11*	*C10orf119*			
RPL3					

APPENDIX C

DNA CONTAMINATION AND RNA QUANTIFICATION

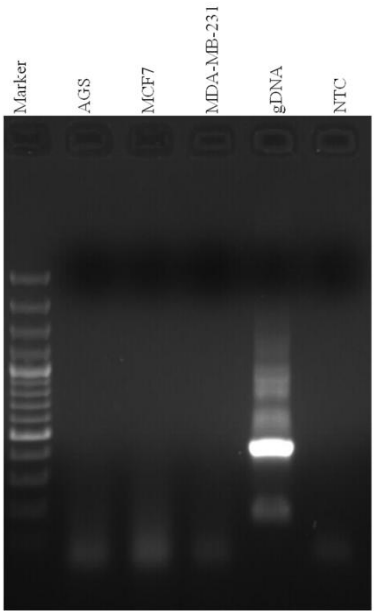


Figure C.1 RNA samples cleared of DNA contamination. Lack of contamination was confirmed by *GAPDH* PCR using the following primers: GAPDH_F: 5'-GGGAGCCAAAAGGGTCATCA-3' and GAPDH_R: 5'-TTTCTAGACGGCAGGTCAGGT-3'. The reaction conditions in 40 cycles were as follows: initial denaturation at 94^{o}C for 2 minutes, denaturation at 94^{o}C for 30 seconds, annealing at 56^{o}C for 30 seconds, extension at 72^{o}C for 30 minutes and final denaturation at 72^{o}C for 10 minutes. Genomic DNA was used as positive control.

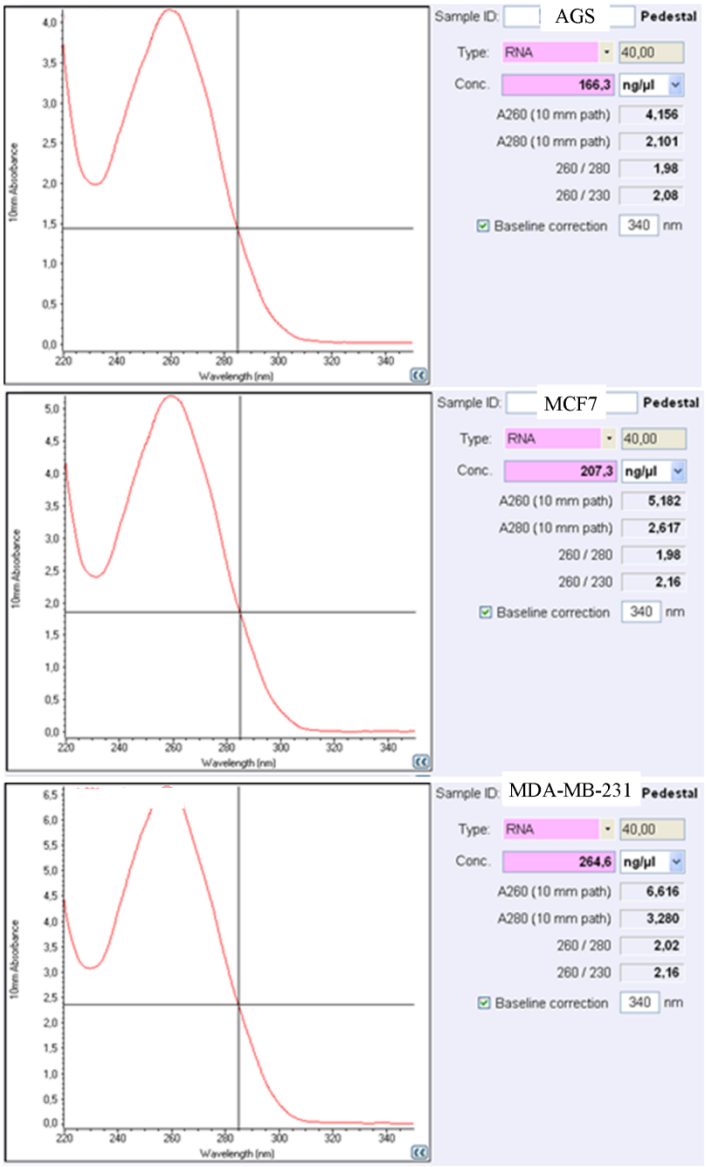


Figure C.2 RNA sample concentrations after DNase I treatment was measured using NanoDrop ND1000 (Thermo Scientific). Sample purities were determined by absorbance ratios.

APPENDIX D

DNA MARKER

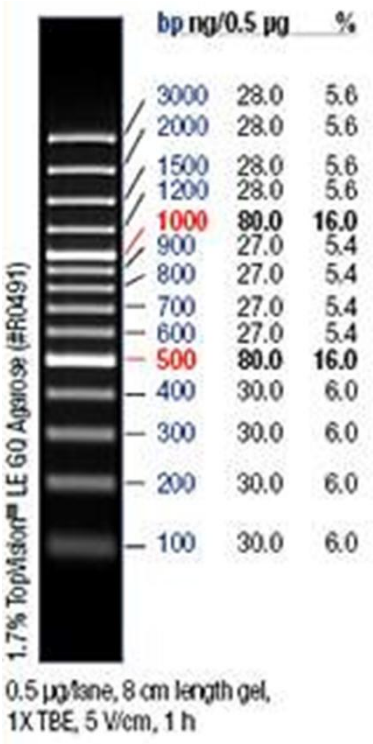


Figure D.1 GeneRuler DNA Ladder 100 bp Plus

APPENDIX E

PLASMID MAP

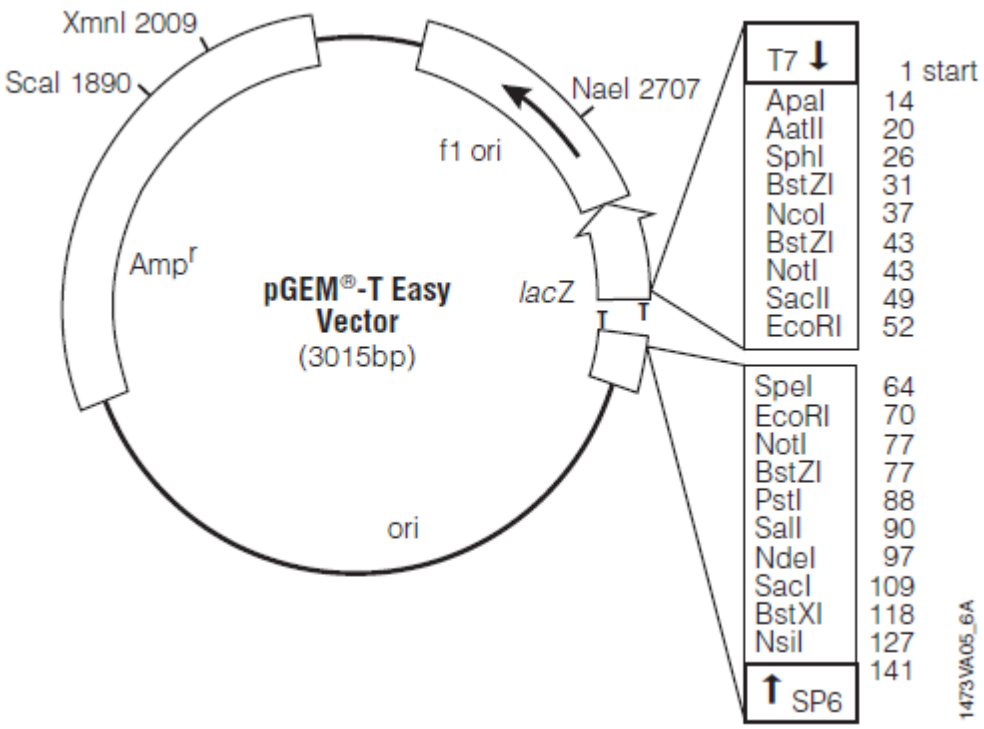


Figure E.1 pGEM-T Easy Vector Map (Promega)