Profiling and Identification of Small Non-coding RNAs as Prognostic Markers for Breast Cancer

By

Preethi Krishnan

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Medical Sciences - Laboratory Medicine and Pathology
University of Alberta

Abstract

Breast cancer (BC) continues to be one of the leading causes of cancer related death among women. Despite continuous progress in screening, diagnosis and treatment of BC, a subset of patients experience recurrence and/or death. Optimal management of BC has remained a challenge due to these inter-individual variations in response to treatment. Although the reasons for inter-individual variations are elusive at this point of time, the challenge now lies in identifying patients who are at higher risk for recurrence and/or death. This in turn may aid in altering treatment modalities according to individual's needs, enhancing the quality of life and survival period. So far, prognostication of BC has relied largely upon clinical staging combined with traditional biomarkers such as Estrogen receptor, Progesterone receptor and human epidermal growth factor receptor but these have remained imperfect estimators of risk for recurrence. Messenger RNA molecules from microarray profiling studies that have so far been in several clinical trials for BC prognostication have also seen limited success in routine clinical use, highlighting the need for more robust biomarkers. In this thesis, I have considered small non-coding RNAs (sncRNAs) as potential biomarkers for BC. sncRNAs (< 200 nt in length) are a group of RNAs that are transcribed, yet not translated, but perform an array of functions. Specifically, I have focused on four sncRNAs – miRNAs, piRNAs, tRNAs and snoRNAs. Although the canonical functions of each of these RNAs are different, these four RNAs appear to share some gene regulatory functions predominantly at the post-transcriptional level, though there may be exceptions for gene regulation even at a transcriptional level. miRNAs and piRNAs are classified as master regulators of gene expression; whereas, tRNAs and snoRNAs are currently being

explored for gene regulatory functions. A possible mechanism by which these molecules may exert regulatory roles is by generating distinct gene regulatory molecules (e.g., miRNAs and piRNAs). The clinical relevance of miRNAs in the context of BC has been well addressed. However, the contribution of the piRNAs, snoRNAs and tRNAs is beginning to emerge for BC etiology but their role in prognosis in BC are at best rudimentary, if not, unknown. The main objective of this thesis was to identify miRNAs, piRNAs, tRNAs and snoRNAs associated with BC prognosis, with outcomes of interest being overall survival (OS) and recurrence free survival (RFS). sncRNAs were profiled from 11 normal (reduction mammoplasty) and 104 breast tumor tissues using next generation sequencing, which enables a genome-wide capture of sncRNAs. Two statistical paradigms were adopted to identify prognostic markers from every class of sncRNAs – case-control (CC) and case-only (CO). While the former approach considered only differentially expressed sncRNAs for survival analysis and may miss on a subset of expressed sncRNAs, the latter approach included all the sncRNAs profiled for a comprehensive analysis. Individual classes of sncRNAs from CC and CO were subjected to Univariate Cox proportional hazards regression modeling. Risk scores were constructed using a panel of significant sncRNAs (which varied from 4-14 for each class of sncRNAs). Based on cut-off point estimated using receiver operating characteristics curve, patients were classified into low and high-risk groups. Further, risk scores were investigated to identify their potential as independent prognostic factors using multivariate Cox proportional hazards regression model. Signatures from miRNAs, piRNAs, snoRNAs and tRNAs independently showed association with both OS and RFS – (i) risk scores were identified as potential independent prognostic factors and (ii)

patients belonging to high-risk group were associated with poor prognosis. sncRNAs associated with OS were independently validated using TCGA dataset, strengthening the study findings. To further gain biological insights of the prognostic sncRNAs, putative gene (mRNA) targets regulated by miRNAs and piRNAs were identified from an in-house gene expression dataset; these studies served as a proxy for functional validation. Also, other sncRNAs (along with their corresponding targets) embedded within snoRNAs were identified. The identified targets were involved in key cellular pathways such as apoptosis, cell cycle, cell migration and proliferation. Overall, my work has identified novel sncRNA molecules as potential biomarkers for BC prognostication. This work on genome-wide profiling of sncRNAs using modern sequencing platforms significantly augments the limited previous literature, and the data provided in this study therefore extends the comprehensive search for BC biomarkers.

Preface

This thesis is an original work by Preethi Krishnan. The research work, conducted as part of this thesis was approved by the local Institutional Research Ethics committee (Health Research Ethics Board of Alberta- Cancer Committee under the protocol #26126).

qRT-PCR validations of the identified small non-coding RNAs were performed collaboratively with Dr. Olga Kovalchuk, University of Lethbridge, Alberta, Canada.

Work from the chapters 3, 4 and 5 have been published in peer-reviewed journals. Individual contributions from all the authors are listed below.

Chapter 3 - Profiling of small non-coding RNAs from human breast tissues – normal and malignant

This chapter was written by Preethi Krishnan. Publications originating from chapters 4, 5, 6 and 7 have common profiling experiments and hence were captured only in this chapter. Therefore contents from this chapter and are organized in a manner to remove redundancy of the material and yet maintain continuity and brevity.

Chapter 4 - Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer

Contents from this chapter appeared as a peer reviewed publication in BMC Genomics. 2015; 16:735 (doi: 10.1186/s12864-015-1899-0).

Author list - Preethi Krishnan, Sunita Ghosh, Bo Wang, Dongping Li, Ashok Narasimhan, Richard Berendt, Kathryn Graham, John R. Mackey, Olga Kovalchuk and Sambasivarao Damaraju.

Author contributions – Preethi Krishnan and Sunita Ghosh performed statistical analysis and interpreted the data; Preethi Krishnan analyzed all of the next generation sequencing data with support provided from Ashok Narasimhan; Olga Kovalchuk, Bo Wang and Dongping Li conducted qRT-PCR experiments and together with Preethi Krishnan offered data interpretations; John R Mackey and Kathryn Graham provided breast cancer gene expression data set; Sambasivarao Damaraju conceived the study, designed the experiments with Preethi Krishnan, provided analysis and interpretations of the data and is the principal investigator of the project; and Richard Berendt provided pathology support. Manuscript preparation was by Preethi Krishnan and Sambasivarao Damaraju; all contributing authors reviewed the manuscript and provided edits and suggestions.

Chapter 5 – Piwi-interacting RNAs and PIWI genes as novel prognostic markers for Breast cancer

Contents from this chapter appeared as a peer reviewed publication in the journal Oncotarget. 2016 (doi: 10.18632/oncotarget.9272).

Author list – Preethi Krishnan, Sunita Ghosh, Kathryn Graham, John R Mackey, Olga Kovalchuk, Sambasivarao Damaraju.

Author contributions – Preethi Krishnan and Sunita Ghosh performed statistical analysis and interpreted the data; Preethi Krishnan analyzed all of the next generation sequencing

data; Olga Kovalchuk, Bo Wang and Mieke Heyns conducted qRT-PCR experiments and together with Preethi Krishnan offered data interpretations; John R Mackey and Kathryn Graham provided breast cancer gene expression data set; Sambasivarao Damaraju conceived the study, designed the experiments with Preethi Krishnan, provided analysis and interpretations of the data and is the principal investigator of the project; the manuscript was prepared by Preethi Krishnan and Sambasivarao Damaraju; all contributing authors reviewed the manuscript and provided edits and suggestions.

Dedicated to

My inspiring husband and my beautiful family

Acknowledgements

First and foremost, I thank my supervisor Prof. Sambasivarao Damaraju for being a fantastic support and an excellent guide right from the day I joined his lab. Thank you for being there always, for having the confidence in me and for all your encouragement. Your three words “You are capable”, has definitely motivated me to work harder. I am thankful for the innumerable lab meetings and discussions we have had over the years. This has definitely helped me gain the much needed confidence. In all, thank you very much for making my Ph.D program, a very pleasant and memorable journey.

I am thankful to my supervisory committee members - Dr. John Mackey, Dr. Olga Kovalchuk and Dr. Sunita Ghosh for their invaluable time, patience and support. Dr. Mackey’s oncology expertise, Dr. Kovalchuk’s expertise on epigenetics and Dr. Ghosh’s expertise on statistics has contributed to my intellectual growth and for successful completion of my project. Overall, it has been a very rewarding experience to have worked with a vibrant interdisciplinary team.

I would like to thank Dr. Wan Lam and Dr. Richard Fahlman for being part of the examining committee and for offering their constructive feedback and comments in my thesis. My sincere thanks to Dr. Michael Weinfeld, for serving as one of the external examiners for my candidacy examination. I am thankful to Dr. Monika Keelan for chairing the examinations and for being a backbone for all the graduate students. I am also thankful to Dr. Carol Cass for her critical reading of my manuscript and for providing constructive feedback.

It has been an enriching experience to have worked with Dr. Bo Wang, Dr. Dong Ping Li, and Dr. Yaroslav Illnytstkyy from University of Lethbridge. I would like to thank Ms. Cheryl Titus, Mr. Ronald Marinas and Ms. Femka Williams for making this a hassle free journey through their administrative support. I would also like to thank Ms. Jennifer Dufour and Ms. Lillian Cook for their technical assistance and for all the help they have offered.

I would like to thank my lab colleague Ms. Mahalakshmi Kumaran for being an excellent support, within and outside the lab. Her critical feedback and stimulating discussions has definitely helped me think outside the box. I will take this experience and wonderful memories for years to come.

My heartfelt thanks to Alberta Cancer Research Biobank/Canadian Breast Cancer Foundation (CBCF) tumor bank for providing access to precious specimens with well-annotated clinical data, without which this study would not have been possible. I am grateful to all the participants of this study. I would also like to acknowledge Canadian Breast Cancer Foundation (CBCF) for providing the funding support for this project.

I am forever indebted to my family: my parents, in-laws and brother for their unconditional love and support. Thank you very much for believing in me, without which this journey would not have been possible. Special thanks for my best friend Ms. Mahalakshmi Sridharan for her indomitable support for all these years. Last but not the least; I thank my husband, best friend and my lab colleague – Ashok Narasimhan, for all his patience, love and support. This travel was made all the more beautiful and enjoyable with you beside me!

Table of Contents

List of tables

List of Figures

List of Abbreviations

AGO	Argonaute
ATM	Ataxia Telangiectasia Mutated
BC	Breast cancer
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
CC	Case-control
CDH11	Cadherin 11
CHEK2	Checkpoint kinase 2
CI	Confidence interval
CO	Case-only
DE	Differential expression
ER	Estrogen receptor
ES	Enrichment score
FC	Fold change
FDR	False discovery rate
FF	Flash/Fresh frozen
FFPE	Formalin fixed paraffin embedded
FISH	Fluorescence in situ hybridization
GEO	Gene Expression Omnibus
GO	Gene Ontology

HER2	Human Epidermal Growth Factor Receptor
HR	Hazard Ratio
IHC	Immunohistochemistry
miRNA	MicroRNA
mRNA	Messenger RNA
NGS	Next Generation Sequencing
OS	Overall Survival
PCA	Principal Component Analysis
PGS	Partek Genomics Suite
piRNA	Piwi-interacting RNA
PIWI	P-element induced wimpy testis in drosophila
PR	Progesterone receptor
PTEN	Phosphatase and tensin homolog
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RFS	Recurrence Free Survival
RISC	RNA Induced Silencing Complex
RNA	Ribonucleic acid
ROC	Receiver operating characteristics curve
RPKM	Reads per kilobase per million
siRNA	Small-interfering RNA
sncRNA	Small non-coding RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA

STK11	Serine/Threonine Kinase 11
TAC	Taxotere/Docetaxel, Adriamycin and Cyclophosphamide
TCGA	The Cancer Genome Atlas
TMA	Tissue Microarray
TNBC	Triple negative breast cancer
TP53	Tumor protein 53
tRF	tRNA derived fragment
tRNA	Transfer RNA
UTR	Untranslated region

1 Introduction

1.1 Breast cancer

1.1.1 Incidence and mortality

Breast cancer (BC) is a malignant tumor that starts in the breast cells lining the ducts or lobules. As the cancer progresses, it may spread to other parts of the body. It is one of the most commonly diagnosed cancers in women worldwide; with approximately 1.7 million new cases being diagnosed. Approximately, 522,000 women die due to BC, making it one of the most common causes of cancer related deaths among women [1]. Among Canadian women, BC accounted to 26% (n = 25,000) of all new cases diagnosed with cancer and represented 14% (n = 5000) of all cancer related deaths in 2015 [2]. There has been an upsurge in the incidence rates of BC, mainly due to increase in awareness and screening programs. Encouragingly, a decrease in the mortality rates is also observed owing to improved therapies [2,3].

1.1.2 Risk factors

Any factor that increases the chance of getting cancer is called a risk factor. Some of the non-modifiable risk factors for BC include age, gender, family history, personal history of BC, race, breast cellular changes, previous exposure to radiation, menarche and menopausal times, pregnancy and breast feeding, breast tissue and bone mineral density [4-9]. Genetic mutations also contribute to the risk of developing BC, especially hereditary (familial) BC [10-12]. The classical risk genes conferring familial BC include *BRCA1* and *BRCA2* which are DNA repair enzymes. Life time risk of BC in inherited *BRCA1*

mutation is 55 – 65% and that of *BRCA2* is 45%. Mutations of other genes such as *ATM, TP53, CHEK2, PTEN, CDH11* and *STK11* may also contribute as risk factors for BC, categorized under familial risk with or without *BRCA* mutations. Some the risk factors that can be controlled (or modifiable) include weight, diet, physical activity, alcohol consumption, use of tobacco, oral contraceptive use, stress and anxiety [13].

1.1.3 Histological subtypes of breast cancer

Most often (50-75%), BC starts in the cells lining the ducts, which are tubes carrying milk from glands to nipple and are classified as ductal carcinoma. Similarly, another class of BC exists, called lobular carcinoma (10-15%) which begins in the lobules (groups of cells that make the glands for producing milk) [13]. Depending on the extent of cell growth, BC can be broadly categorized into two classes: Carcinoma in situ (cancer is localized and has not grown into the surrounding tissues) and invasive breast cancer (cancer has grown into the surrounding tissues). Both ductal and lobular cancers can be either in situ or invasive. In-situ cancers are often curable but the life-time risk of breast cancer occurrence is high. Approximately 90% of all cancers diagnosed belong to the invasive type. Several "special" types of invasive carcinomas occur such as tubular carcinoma, medullary carcinoma, mucinous/colloidal carcinoma, papillary carcinoma and cribriform carcinoma. These are less frequent when compared to the invasive ductal and lobular carcinomas. Some of the other rare histological subtypes of BC include inflammatory breast cancer, paget's disease of the nipple and phyllodes tumor of the breast [14].

1.1.4 Molecular subtype classification of breast cancer

We now understand that breast cancer is not a single disease. Pioneering work by Sorlie et al. revealed four main subtypes of BC based on gene expression profiling using microarrays. According to this classification, BC can be classified into Luminal, which can be further divided into Luminal A and Luminal B, HER2+ enriched and Triple negative breast cancer (TNBC) [15-17] based on the expression patterns of hormonal receptors – estrogen receptor (ER-□; here after referred to as ER), progesterone receptor (PR) and human epidermal growth factor receptor (HER2). BC classification has evolved further over time and we now have further subtypes based on gene expressions profiles (e.g. Claudin-low). BC classifications have helped to understand the survival patterns of each of these subtypes, their prevalence and tumor characteristics unique to the individual subtypes; suggesting that each of these subtypes may behave as different disease entities and this adds further complexity for optimal management of BC [17-21]. Among these subtypes, Luminal A is a good prognosis subtype of BC relative to other BC subtypes. However, it is not uncommon to see patterns of late recurrences (local or distant metastatic spread) in this subtype, suggesting the need for further research to reduce the mortality associated with this BC type. On the other end of the spectrum, we have the TNBC subtype comprising 10-15% of all BCs, is associated with extremely poor prognosis. The Luminal B and HER2+ enriched subtypes show intermediate prognosis. My thesis focus is restricted to identify prognostic markers for Luminal A and TNBC, as these represent majority of invasive BCs with extreme ends of the prognoses spectrum. Molecular subtypes of BC, along with their characteristics are summarized in Table 1.1.

Table 1.1 Molecular classification of breast cancer

Molecular subtype	Prevalence	Marker expression profile	Representative Characteristics [15-21]
Luminal A [15-21]	30-70%	ER+/PR+/HER2- ; Any type of cytokeratin(CK) ; Ki67 – low	Low proliferation rate; Low grade; Good prognosis; Low relapse; High survival rate; Metastasis to bone, CNS, liver, lung;
Luminal B [16-21] **(Luminal HER2)**	10-20%	ER+/PR+/HER2- ; Ki67 – high ; Cyclin B – high ER+/PR+/HER2+ ; Ki67 – low or high	More aggressive; Intermediate/higher tumor grade; High proliferation rate; Intermediate/Worse prognosis; Fairly high survival rate but not as high as Luminal A; Intermediate p53 mutations
HER2+ (HER2+ enriched) [16-21]	5-15%	ER-/PR-/HER2+	Aggressive; High proliferation rate; High tumor grade; Worse prognosis; Recurrence rate high; Metastasis frequent; High p53 mutations
Basal [15-21] **Triple Negative breast cancer**	15-20%	ER-/PR-/HER2- ; EGFR+ or CK5/6+ ER-/PR-/HER2- ; EGFR-/CK-	High tumor size; High tumor grade; High frequency of lymph node involvement; More aggressive; High recurrence rate (most likely within 3 yrs); High mitotic index High p53 mutations Poor prognosis
Normal breast like [16-21]	5-10%	ER-/PR-/HER2- ; CK5-/EGFR-	Intermediate prognosis between luminal and basal; Do not respond to neo-adjuvant tx; Very rare; Low proliferation; Low grade; Low p53 mutations
Claudin-low [18-20]	12-14%	Low expression of genes involved in tight junctions and intercellular adhesion including claudin 3,4,7, occludin, E-cadherin ; ER-/PR-/HER2-	Over expresses a set of 40 genes related to immune response; Poor prognosis; Overexpress genes linked to mesenchymal differentiation and EMT (associated with acquisition of stem cell properties)

1.1.5 Prognostic and predictive factors for breast cancer

Biomarkers to guide treatment decisions are broadly classified as either prognostic or predictive factors. Prognosis is the estimate of the likely outcome of a disease. A prognostic factor is defined as "a clinical or biologic characteristic that is objectively measureable and that provides information on the likely outcome of the cancer disease in an untreated individual" [22]. Prognostic factors are therefore helpful in identifying patients who are at risk for recurrence and/or death, which may eventually help in modifying treatment modalities. On the other hand, predictive factor is defined as "a clinical or biologic characteristic that provides information on the likely benefit from treatment (either in terms of tumor shrinkage or survival)" [22]. Predictive factors may therefore aid in identifying patients who are likely to respond to a treatment. Some of the factors that determines the prognosis of breast cancer include tumor size, lymph node status, tumor stage, tumor grade, age, tumor type, receptor status, HER2 status, subtypes based on gene expression profiling, proliferation rate, menopausal status, general health and tumor recurrence [23-28]. Some of these factors are summarized in table 1.2.

Table 1.2 Factors determining breast cancer prognosis

Factor	Features assessed	Inferences
Tumor size	Overall size of the tumor	Generally, higher the size, poorer is the prognosis; Size helps in determining the tumor stage
Lymph node status	Number of nodes with tumor cells	Nodes with tumor cell infiltration are associated with poor prognosis; Lymph node status is important for determining the stage
Tumor stage	Tumor size, lymph node status and metastasis	Higher stage is associated with poor prognosis
Tumor grade	Cell morphology, division and tubule formation	Higher grade tumors are associated with poor prognosis
Tumor type	Histological and molecular subtypes	Invasive cancers are associated with poor prognosis; Basal type and HER2+ enriched tumors are associated with worse prognosis, compared to Luminal types
Hormonal receptor status	Estrogen receptor, progesterone receptor	Positive expression of receptors tend to be associated with good prognosis; Receptor positive tumors can be treated with hormonal therapies; ER and PR can act as prognostic and predictive factors
HER2 status	Amount of HER2 protein expressed in the surface of cells	Overexpression of HER2 associated with poor prognosis; HER2+ tumor patients respond to trastuzumab; HER2 can be used as prognostic and predictive factor
Age	Age at the time of diagnosis	Younger age is associated with poorer overall prognosis

BC prognostication relies largely upon the above mentioned traditional factors. However, these molecules have limited efficacy in accurately predicting the recurrence risk and have necessitated the identification of newer molecules for prognosis. For instance, lymph node status has been a good indicator of prognosis. Yet, 30% of patients with node negative tumor develop recurrences within 10 years [29]. Similarly, tumor grade is a qualitative assessment and comparative analysis between three independent pathologists has shown less than 50% concordance [30], indicating that the accuracy of risk estimates may vary considerably depending upon the individual who handles the samples.

Biomarkers currently in use (ER, PR and HER2) are highly informative but these markers are also of limited utility to predict individual outcomes [31]. ER+ status in general indicates good prognosis but Luminal A tumors have the potential for late recurrences [32], (>10 years). Likewise, HER2+ and PR are also good prognosticators but are imprecise estimators of distant recurrences [30], warranting the need for other prognostic markers of higher specificity and sensitivity. Availability of such markers as stand-alone markers or when used in combination with the currently used ER, PR and HER2 markers may offer to guide treatment decisions favoring better outcomes. ER, PR and HER2 markers are unique in that they serve as both prognostic and predictive markers.

1.1.6 Gene expression signatures for breast cancer prognostication

Recent advances in global gene expression profiling using microarray or next generation sequencing platforms have informed our understanding of pathways contributing to BC etiology [33,34]. Based on this premise of unbiased profiling of global gene expression, several research groups had taken upon the task of developing multigene markers for BC prognosis. Further, developments in Tissue Micro Arrays (TMAs) aided in low- medium throughput profiling of a number of protein markers based on immunohistochemistry (IHC), or DNA/RNA based Fluorescence in situ hybridization (FISH). Once a gene signature is established, use of RT-PCR and custom microarray platforms helped translate these findings to routine clinical use. Each of the tests developed using these platforms have advantages and disadvantages and are summarized below [35,36].

1.1.6.1 IHC based multigene predictors

Two main assays were developed using IHC – ProEx Br and Mammostrat [37]. Both the tests use five antibodies. Overexpression of two or more of these markers have been found to be associated with relapse in ProEx Br. Mammostrat is a commercially available test (Applied Genomics Inc., Huntsville, AL) that estimates the risk of recurrence in ER+, lymph node negative tamoxifen treated patients based on slide scoring of five antibodies. Patients are then classified into low-, intermediate- and high-risk groups. IHC platforms offer the advantages of low false discovery, less statistical algorithms needed for data analysis and a comparatively lower cost for profiling. However it is faced with other challenges [38], such as variations involved in tissue processing (Formalin Fixed and Paraffin Embedded or FFPE), quantitative scoring of immunohistochemical staining, estimating the optimal cut-offs for identifying risk groups; all of which can have a significant impact on determining the prognosis of a patient. Nevertheless, the stand alone prognostic value of IHC is well established in breast cancer with the routine use of ER, PR, HER2 and Ki67 as prognostic and/or predictive markers.

1.1.6.2 FISH based multigene predictors

FISH based testing is predominantly used to estimate the expression levels of HER2 gene. A multicolor FISH assay (to estimate copy numbers of three genes) has also been developed by eXagenBC (eXagen Diagnostics, Inc., Albuquerque, NM) [39] as a pure prognostic assay for both node positive and negative ER+ patients. FISH assay is also influenced by pre-analytical variables and encounters other technical challenges as in IHC when using FFPE specimens.

1.1.6.3 RT-PCR based multigene predictors

Among the multigene assays, Oncotype Dx (Genomic Health, Inc., Redwood City, CA) is one of the top assays considered for clinical decision making in BC. It is a 21 gene (16 informative and 5 reference genes) prognostic and predictive assay based on RNA extracted from FFPE samples [30]. A recurrence score is computed based on these 21 genes and patients are classified into three risk groups: low, intermediate and high-risk groups. This test was developed mainly for ER+ lymph node negative patients. However, this study is slowly showing promise for lymph node positive patients as well. Since the time of development of this assay, there has been a change in the classification of patients based on the tripartite recurrence score. Oncotype Dx has entered into clinical trials (TAILORx) and recent results from the trials provide prospective evidence that this multigene prediction assay can be used to identify patients at low risk for recurrence who can be spared from chemotherapy [40]. One of the major drawbacks of this test is that so far, the benefit of adjuvant chemotherapy for patients belonging to the intermediate risk group is not clear [41]. Another challenge with Oncotype Dx is the estimation of HER2 levels. IHC and FISH techniques are routinely used to estimate HER2 expression but some studies have demonstrated discordance in HER2 expression levels between RT-PCR and the traditionally used assays (IHC and FISH) [42]. Since this assay relies heavily on HER2 expression, the recurrence score based classification of patients may be questionable. An independent study was conducted by Cuzick et al to compare the recurrence score with prognostic score estimated using the traditional markers – ER, PR, HER2 and Ki67. Prognostic information provided by both the sets of markers was

similar, questioning the usefulness of this assay over the traditional markers, which are simple to assess [43].

Another important test that is commercialized is PAM50 gene signature [44], developed by Parker et al. This assay was mainly developed for standardizing subtype classification of breast cancer and the identified gene signature also showed prognostic benefit.

Other RT-PCR based assays including Breast Cancer Two Gene Expression Ratio H/I™ (Aviara Dx, Inc., Carlsbad, CA, USA) [45], the Celera Metastasis Score™ (Celera, Inc., Rockville, MD, USA) [46] and the Breast BioClassifier (Associates in Regional and University Pathologists, Salt Lake City, UT, USA) [47] are also available, but none have reached the stage of prospective validation and are still not available for clinical application.

1.1.6.4 Microarray based multigene predictors

Prognostic assays developed using microarray platforms have typically used fresh frozen tissue samples. In all these assays, it is vital to critically analyse the samples for the presence of any normal cells as the number and expression of RNAs identified depends largely on the composition of the sample (tumor and normal cells). One of the important and the first fully commercialized microarray based assays is Mammaprint (Agendia BV, Amsterdam, The Netherlands), developed for ER+ or ER- lymph node negative patients under the age of 61 [48]. This is a 70 gene signature assay that is most useful to identify extremes of disease outcome (low risk and high risk). The assay was subsequently validated in an independent study and the 70 gene expression signature

stood out as the strongest predictor for metastasis free survival [49]. However, the strongest criticism faced by this assay is that the validation set also included samples from the discovery set, leading to overestimation [50,51]. MINDACT trial is a prospective study that assesses patients based Adjuvant! online as well as the 70 gene expression signature [52]. Patients who were classified as belonging to low-risk and high-risk in both the tests were recommended adjuvant chemotherapy and endocrine therapy, respectively. However, treatment for patients with discordant results from both the tests was either adjuvant chemotherapy in addition to endocrine therapy or endocrine therapy alone. The whole objective of this trial is to identify patients with low risk to avoid overtreatment.

Following Mammaprint, other microarray based assays have also been developed, such as Rotterdam signature (also called the 76 gene assay) [53], invasiveness gene signature [54], Nuvoselect assay [55,56], among others.

With improving technological platforms, there has been a surge of biomarkers for BC prognosis. Yet, the common practice is to assess the expressions of ER, PR and HER2, since these molecules show both prognostic and predictive behavior. Different assays have adopted different approaches to develop a multigene signature for BC prognosis and various factors contribute to the identification of the best set of markers with prognostic benefit. For instance, some assays have focused on proliferating genes and other genes playing a role in hallmarks of cancer. Profiling these signatures require careful assessment of the tumor sample obtained so as to include only tumor cells and not the stromal cells. The type of samples obtained i.e. fresh frozen or FFPE samples have a significant influence on the type of signatures obtained. While the probability of

obtaining intact RNA is higher in fresh frozen samples, yet obtaining and maintaining such samples are not cost-effective. On the contrary, FFPE samples are readily available and are more useful to run a retrospective analysis with long follow-up periods. However, the quality of mRNA obtained from such specimens in handling of fresh tissues is critical for overall success of the assays [57]. For these reasons, some of the commercially available assays are now testing the feasibility of the developed signature in FFPE samples. As explained above, most of the assays have been developed for ER+ and lymph node negative tumors. The risk assessment for other types of cancer is still in question. Even though ER+ tumors are considered to be good prognosis tumors, the chances of late recurrences is higher in this subtype. Therefore, when an assay is developed, it is critical to include samples with longer follow-up periods. The main concerns regarding the assays already developed arise over their scientific validity, true clinical utility, cost/benefit ratios and their restriction to specific clinical settings. Currently, we also do not know if these assays perform better as stand-alone markers or if they complement to the traditional markers. If there is no benefit over the routinely used markers, the utility of the newly developed markers is debatable. While the validation process and the trials of the developed assays continue to make strides, we also need to identify other biomarkers that may overcome the shortcomings of the developed assays and contribute to better prognostication or prediction than the currently developed ones.

1.1.7 The need for prognostic markers

Over the past three decades, there has been an increase in the age adjusted incidence rate of BC patients in the US [58]. However, if we consider the last ten years, the incidence rate has been stable, owing to increased awareness and screening measures.

There has also been a steady decline in the death rates, suggesting improvements in BC therapies. For instance, the five year relative survival rate was estimated to be 91% in 2007 [58]. Common clinical practice for BC treatment includes tumor resection, followed by adjuvant systemic chemotherapy, endocrine therapies and/or radiotherapies. While these therapies are beneficial, they are not free from drug-associated toxicities [59]. However, while some patients tolerate treatments and respond better, others develop toxic effects. Therefore there is an unmet need to accurately identify patients who may benefit more from treatment from patients who may not benefit and spare them from unnecessary treatment. Despite improved adjuvant therapies and improved survival rates, about 20-30% of BC patients develop metastasis [18,20,60], which at this point remains incurable, leading to unfavorable outcomes. Therefore it is critical to identify patients who are most likely to develop recurrence and/or die. This stratification of patients based on their risk for recurrence and/or death may help in developing tailored therapies and enable further improvement in the survival rates. As explained earlier, several prognostic and predictive markers are available but three (ER, PR and HER2) are routinely used. Nevertheless, the need for biomarkers continues to exist as the traditional markers remain as imperfect estimators of risk for recurrence.

1.2 Non-coding RNAs

For a long time, the field of molecular biology has been governed by the central dogma, which can simply be explained as DNA makes RNA, and RNA makes protein. While this still holds good, recent discoveries have subverted this principle. A group of RNAs termed “non-coding RNAs” have been found to play a role in regulating

transcription and translation. These RNAs were previously considered not to play significant roles in human system but we now understand that they are involved in diverse roles ranging from gene regulation to alternative splicing to protein translation. Several classes of non-coding RNAs have been discovered but these are broadly classified into two groups based on their size: long non-coding RNAs (lncRNAs), which are generally > 200 nucleotides and small non-coding RNAs (sncRNAs), which are generally less than 200 nucleotides. One of the major functions of these two groups of RNAs lies in regulating gene expression. An important characteristic of lncRNAs is that they contain exons and introns as in mRNA/protein coding genes. As such, lncRNAs resemble protein coding genes in terms of several sncRNAs embedded within. The focus of this thesis is on sncRNAs. SncRNAs include several classes such as microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs) and small interfering RNAs (siRNAs).

1.2.1 microRNAs (miRNAs)

MicroRNAs are small (~22nt), non-coding, regulatory RNAs that control gene expression post-transcriptionally by binding to the 3'UTR of mRNA and promote mRNA degradation or inhibit protein translation [61-64].

1.2.1.1 Discovery of miRNAs

In 1993, Victor Ambros, Rosalind Lee and Rhonda Feinbaum discovered that lin-4 , a heterochronic gene involved in the temporal developmental pattern of *C.elegans* did not code for any protein, instead produced a pair of small RNAs [61]. The two small RNAs were approximately 22 and 61 nucleotide in length and had sequences

complementary to the 3'untranslated region (UTR) of lin-14 mRNA [61]. At the same time, Gary Ruvkun and his colleagues – Bruce Wightman and Ilho Ha discovered that lin-14 was post transcriptionally regulated by lin-4: lin-4 base paired with the 3'UTR of lin-14 leading to down regulation of lin-14 translation [62]. The smaller length RNA (~22nt long) discovered by Lee et al, is the first member of microRNA family of small RNAs [61,63]. It was only after seven years in the year 2000 that the second miRNA (let-7) was discovered by Reinhart et al [65]. The fact that let-7 was conserved across species revolutionized the research on small RNAs. As of April 2016, 2,588 unique mature human miRNAs have been identified on the human genome [66,67].

1.2.1.2 Location of miRNAs in human genome

miRNAs may be identified from the intergenic regions, or from exonic regions or intronic sequences of protein coding and non-protein coding transcriptional units [64,68-71]. Further, some miRNAs may be in a distant location from other miRNAs, while some others may be in proximity and may exist as clusters. A cluster, as defined by miRBase (http://www.mirbase.org/), is a group of miRNAs are located within 10kb of each other [72]. miRNAs belonging to the same cluster may either be co-transcribed or transcribed independently [73-75]

1.2.1.3 Biogenesis of miRNAs

The biogenesis of miRNA begins in the nucleus where miRNA genes are transcribed by RNA polymerase II or III into several kilobases long primary transcripts (pri-miRNAs) that are polyadenylated at the 3'end and capped at 5' end [76-78]. Pri-miRNAs contain stem loop structures and are cleaved at the stem of the hairpin structure

by cellular RNase class II endonuclease III enzyme called Drosha along with DGCR8/Pasha into hairpin structures called precursor-miRNAs (pre-miRNAs) which are approximately 70-120 nt long [77]. The pre-miRNA harbors a 5'- phosphate and a 2 nucleotide overhang, characteristic of endonuclease III enzyme. pre-miRNA is then transported to the cytoplasm with the help of Exportin 5 along with Ran–GTP which is then processed by the cytoplasmic dsRNase III Dicer into approximately 22 nt miRNA: miRNA duplex with 2 nt overhanging at its 3' end [64,68]. The duplex is unwound by helicase and only one mature strand (~20nt long) enters the multicomponent complex called RNA Induced Silencing Complex (RISC), which harbors argonaute proteins (Ago), and the other complementary strand is degraded [64]. Mostly, the strand with relatively unstable base pairs at the 5'end survives [79,80]. The mature miRNA mediates gene expression regulation by binding to the complementary sequence in the 3' untranslated region of target messenger RNA (mRNA). Depending on the complementarity shared, the target mRNAs may be degraded (if the two RNAs are perfectly complementary to each other) or the protein translation may be inhibited (if they share imperfect complementarity) [81].

1.2.1.4 Mechanisms of action

The interaction between miRNA and target mRNAs predominantly occur at the seed region (2-8 nt in the 5'end of the miRNA) of miRNA and the 3'UTR of mRNA through sequence complementarity [82]. This interaction can have several consequences [81], as outlined below. Two main effects have been observed – direct and indirect effects on translation. In the direct effects, initiation of translation or post-initiation of translation is inhibited. While in the former, the association of ribosome with target mRNA is

prevented, the latter includes premature ribosome fall off, reduced/stalled elongation or co-translational protein degradation. Indirect effects of miRNA-mRNA interaction include deadenylation, resulting in degradation or increased turnover. These effects occur in the cytoplasm, predominantly in the processing bodies (P-bodies), which are enriched for factors involved in mRNA degradation. The mRNAs whose protein formation is prevented (by direct or indirect effects) may be sequestered in the P-bodies, which can be used later for translation or can be degraded.

1.2.1.5 Different facets of miRNAs

Since their discovery in 1993, miRNAs have been studied in great depth [83] for their role as key players in normal developmental processes [84] including cell growth and apoptosis [85,86], hematopoietic lineage differentiation [87], muscle cell proliferation and differentiation [88]; tumorigenesis [89,90] and other diseased states such as cardiovascular disease [91,92], autoimmune diseases [93,94] and neurodegenerative diseases [95,96]. The evidence that miRNAs are deregulated in cancers was first observed and demonstrated by Calin et al., in chronic lymphocytic leukemia in the year 2002 [97]. They observed deletion at 13q14 locus in chronic lymphocytic leukemia, which also harbored miR-15/16 cluster. This was not only down-regulated in cancer but was found to regulate BCL2. Since then, several miRNAs have been reported as tumor suppressors [98-100] and oncogenes [101-104]. It is now known that miRNAs regulate approximately 60% of the protein coding genes [84]. Apart from their potential to distinguish normal samples from tumor samples, they have also been valuable biomarkers in clinical diagnostics to help trace the origin of cancer in disseminated conditions [105]. The relevance of miRNA profiling in cancer was established when miRNA profiles accurately reflected the developmental lineage and differentiation

state of tumors, in contrast to the inaccurate results obtained from the mRNA profiles [89]. Although the field of miRNAs is hardly 20 years old, yet it has seen a tremendous progress to an extent that miR-122 is already in phase III clinical trials for Hepatitis C virus infection [106]. Other miRNAs such as miR-34a has reached phase I clinical trial for Liver cancer [100].

The role of miRNAs as biomarkers has been widely studied in several cancer types and its significance as prognostic [107-110], diagnostic [111,112] and predictive markers [113,114] is well established. miRNAs have also been extensively studied for breast cancer as promising biomarkers [115-118]. However, we have still not been able to obtain a consensus miRNA signature for BC prognosis as the science of identifying prognostic markers is ever expanding. This extensive research on miRNAs was the cornerstone to conduct this thesis. Even though miRNAs have been identified as prognostic factors, there is paucity of literature in comprehensive and whole genome mining of miRNA signatures for prognostication. Recent annotations of miRNAs on the human genome and availability of next generation sequencing (NGS) platforms for whole genome capture of miRNAs stimulated my interest in independently replicating previous findings, and in the potential to identify additional and novel miRNAs. NGS allows unbiased profiling of all miRNAs, which are otherwise limited on array (hybridization) based methods.

1.2.2 Piwi-interacting RNAs (piRNAs)

piRNAs are a recently discovered (2006) class of small non-coding regulatory RNAs that are slightly longer (25-32 nt) than the miRNAs and whose role was believed to lie predominantly in germline maintenance and development. However, recent studies

indicate additional roles that piRNAs may play in somatic tissues as well. Similar to miRNAs, piRNAs also interact with Ago proteins to guide target specific gene regulation. Two classes of Ago proteins exist: AGO and PIWI (P-element induced Wimpy testis). While the former class more often interacts with miRNAs, the latter class is associated with piRNAs. Four human homologs of AGO [AGO1, AGO2, AGO3, AGO4] proteins and four of PIWI class of proteins are described. These are HIWI (PIWIL1), HILI (PIWIL2), PIWIL3, HIWI2 (PIWIL4) proteins [119].

1.2.2.1 Discovery of piRNAs

In the year 2006, piRNAs were isolated from mouse testis independently by four groups [120-123]. These RNAs were found to be more abundant than the other small RNAs and it was estimated that every spermatid would approximately, contain at least one million piRNAs. The mouse specific PIWI proteins were found to be expressed in a temporal manner and these proteins were found to interact with small single stranded molecules [124,125]. Based on the length, two classes of RNAs were found to be interacting with PIWI proteins: the length of one class of RNAs ranged from 26-28 nt [121] and the second class had a size range of 29-32 nt [120]. Initially they were believed to be part of repeat associated siRNAs as they showed regulation of repetitive elements such as transposons in the germline. With the observation that they were integrated with PIWI proteins and that they did not require Dicer for their biogenesis, distinguished them from both miRNAs and siRNAs and were thus named as piwi-interacting RNAs or piRNAs [126].

1.2.2.2 Genomic location of piRNAs

While one of the characteristic features of miRNAs is its evolutionary conservation across species, piRNAs do not share sequence conservation [120,121] except for a uridine bias at the first base [127]. In drosophila, piRNA clusters are mostly seen in the repetitive sequences and in regions devoid of protein coding genes. In contrast to miRNAs, piRNAs arise from two genomic sources: piRNA clusters, which is the main source and from protein coding genes [126,128]. Two types of piRNA clusters have been identified, depending on the direction of transcription: unidirectional and bidirectional. piRNAs arising from the second source were initially observed to arise from 3'UTR of protein coding genes [128]. However, a recent study by Martinez et al., have observed mapping of piRNAs to intronic regions of protein coding and non-protein coding genes (e.g., long non-coding RNAs) [129].

1.2.2.3 Biogenesis of piRNAs

The biogenesis pathway of piRNA remains elusive and majority of our understanding stems from our knowledge on Drosophila pathway. Often, piRNA biogenesis is also associated with silencing of target genes. piRNAs take two routes for their processing: primary synthesis pathway and the secondary pathway/ping-pong amplification [130-133]. It is believed that the primary biogenesis pathway is necessary to initiate PIWI pathways, while the secondary pathway is necessary for both, maintaining the piRNA levels and for target silencing. Primary synthesis begins with the transcription from piRNA clusters by RNA polymerase II. After further processing of these sequences, the piRNAs pair with PIWI proteins, and the pair may subsequently re-enter into the nucleus and silence transcription of a target gene. Several proteins such as Zucchini,

Aubergine (Aub), etc., have been found to play a role in the primary pathway of drosophila but no conclusive evidences have been drawn yet. In the somatic cells, we only observe primary pathway, whereas in the germ cells, we observe both primary and secondary pathways [134].

The secondary mechanism, known as the amplification cycle involves only Aub and Ago3 and not PIWI proteins. piRNAs generated from the primary pathway may enter into the secondary pathway and subsequently bind with Aub. In this cycle, the binding of piRNAs to Aub and Ago3 alternate with each other and the sequences that bind to these proteins are complementary to each other. Briefly, the piRNA-Aub complex binds to a target RNA, cleaves it and generates a new sequence, which then binds with Ago3. piRNA-Ago3 complex performs similar mechanism of cleaving the target RNA and simultaneously generating the piRNA.

1.2.2.4 Mechanisms of action

Similar to the mechanism of miRNAs, piRNAs also associate with RISC complex, forming piRISC and protects the genome by silencing transposons. piRISC can also be effective in gene silencing, similar to miRNAs [135]. One of the recent discoveries has also suggested deadenylation of mRNA by piRNA in Drosophila embryos [136]. There are other functions of piRNA-PIWI complexes but their mechanisms of action remains abstract.

1.2.2.5 Functions of PIWI proteins and piRNAs

Functions of PIWI proteins can be classified under two categories – Developmental functions and Regulatory functions [132]. The development functions can further be categorized into germline and somatic functions.

1.2.2.5.1 Developmental functions

Yet again, our understanding on the development functions of PIWI proteins originates from Drosophila, mice, *C.elegans* and other lower order organisms. The major roles of PIWI proteins in germline function include the formation of germ cell, maintenance of germline stem cells, meiosis, spermiogenesis and oogenesis. Gene knock-out and knock-in experiments have revealed the contribution of PIWI proteins in these functions. The significance of PIWI proteins has expanded beyond germ cells to somatic tissues. For instance, they are known to mediate epigenetic regulation and stem cell maintenance in Drosophila, maintenance of neoblast cells in Planaria [137,138]. The development of ciliates involves germline micronucleus and somatic macronucleus. Certain amount of DNA sequences found in the somatic macronucleus has to be eliminated during sexual reproduction and PIWI proteins are known to play a major role in DNA elimination [139]. Knowing the relationship between cancer cells and stem cells, it is not surprising to see the dysregulation of PIWI proteins in human cancers [140], indicating that they may also likely contribute to tumorigenesis.

1.2.2.5.2 Regulatory functions

PIWI proteins may serve as epigenetic suppressors or activators, depending on the recruitment of certain proteins [141,142]. It has also been noticed that transposon coding

genes are not methylated in the absence of PIWI proteins, reflecting to a loss of epigenetic control [143]. piRNAs and PIWI proteins may both serve as upstream mediators of epigenetic control and may also be involved in transcriptional gene silencing [144,145]. The role of piRNAs and PIWI proteins in silencing transposonal activities is well studied. It is believed that piRNAs occur as cluster, especially from the repetitive elements [146]. A specific example would be the flamenco region in flies, which harbors one of the largest piRNA cluster. A disruption in the flamenco region interrupts with the production of piRNAs, with a simultaneous increase in transposon activity [127]. Also, the biogenesis pathway of piRNAs also serves dual purpose – to generate piRNAs and to mediate gene silencing. Co-fractionation of PIWI proteins and piRNAs with polysomes has hinted at the possibility of a potential role for PIWI proteins and piRNAs in translational control [122]. One of the important observations is the role of piRNAs in post-transcriptional gene silencing. Although the mechanism still remains unclear, it is believed that piRNAs may act in a manner similar to that of miRNAs. A study from Esposito et al. opened up newer avenues for exploration in this domain [147]. piR_015520 was found to negatively regulate its host gene MTNR1A gene, offering new functions for piRNAs, similar to miRNAs. Other studies have also confirmed the relationship between piRNAs and its corresponding target mRNAs [148,149], even though it is not known if the piRNAs have any seed sequence that determines its complementary binding with the target mRNA.

1.2.2.6 Role of PIWI proteins and piRNAs in cancer development

PIWI proteins and piRNAs are new players in tumorigenesis. The first report to suggest the role of PIWI proteins in cancer originated from the study on seminomas by Qiao et al [140]. Extending on this study, Lee et al also observed phenotypic differences

relating to induction of PIWI protein expression [150]. These pioneering reports were followed by other studies that focused on the understanding of the contribution of PIWI proteins and piRNAs to cancer. Dysregulation of PIWI proteins have been detected in breast cancer, cervical cancer and have been linked to cell proliferation, apoptosis, invasion and metastasis [151-154]. Clinical significance of PIWI proteins have also been reported. PIWI proteins have shown to possess prognostic significance for gliomas [155], pancreatic cancer [156], colorectal carcinoma [157], to name a few. The expression patterns of PIWI proteins were found to be different in different BC stages [152], indicating their potential to be a biomarker but no study has yet highlighted the prognostic significance of PIWI proteins for BC. Likewise, piRNAs have also been observed to be dysregulated (serving as tumor suppressors or oncogenes) and thus influencing phenotypic effects in different cancer types such as breast cancer [158] and bladder cancer [149], to name a few. Given their diverse roles, the role of piRNAs as biomarkers has also been investigated. Although literature is scanty in this regard, potential of piRNAs to serve as prognostic and diagnostic biomarkers is high [129,159,160]. A thorough study on piRNAs as prognostic markers for breast cancer is still lacking.

1.2.3 Transfer RNAs (tRNAs)

tRNAs are a class of small non-coding RNAs which are 75-95 nucleotides in length[161] and are well known for their role in protein synthesis. A total of 625 tRNA genes have been annotated so far in the human genome, of which 506 are tRNAs that decode standard amino acids, three are selenocysteine tRNAs, three are suppressor tRNAs, three are tRNAs with undetermined or unknown isotypes and 110 are tRNAs predicted to be pseudogenes[162]. The striking feature of a tRNA molecule is its complex

clover shaped secondary structure that is made of three hairpin loops and one terminal helical stem. A tRNA interacts with a messenger RNA at the anticodon loop and carries the corresponding amino acid at the 3' end. Since the genetic code is degenerate (one amino acid can be encoded by more than one codon), it is implied that tRNAs will exist for every codon that codes for an amino acid and these tRNAs are called as isoacceptors but the wobble hypothesis reduces the number of tRNAs needed and in all, 46 tRNAs are sufficient for 61 codons [163,164].

1.2.3.1 Genomic location of tRNAs

Despite being one of the oldest molecules discovered so far, we do not precisely know the exact genomic locations from which the tRNAs arise. In some organisms such as trypanosomes, tRNAs are present in the boundaries of transcriptional units [165]. In some instances, tRNAs have been identified as clusters (distance between two tRNA genes is less than 1000 nucleotides) [166].

1.2.3.2 Discovery and biogenesis of tRNAs

The discovery of tRNAs dates back to 1956 when Paul Zamecnik and Mahlon Hoagland identified an adaptor molecule that functioned as an intermediate carrier of amino acids in protein synthesis [167,168].

The biogenesis of tRNAs begins with the transcription of a tRNA gene by RNA polymerase III and with the help of transcription factors. The initial precursor tRNA is subjected to trimming of the 5' leader and 3' trailer sequences , followed by the addition of CCA sequence at the 3' end. This sequence is important for the attachment of amino acids to tRNAs. Often, tRNA genes embed intronic sequences which needs to be spliced

out. tRNAs also undergo several modifications to become a structurally and functionally stable molecule. In all these steps, they are termed as uncharged tRNAs and with the attachment of aminoacylated tRNAs, they are termed as charged tRNA molecules or aminoacylated tRNAs and this completes the formation of a mature tRNA. Only the aminoacylated tRNAs can take part in protein synthesis [169,170].

1.2.3.3 Functions of tRNAs

The well-established and well characterized function of tRNAs is its role in protein synthesis. Over the years, their roles have expanded beyond their canonical function in translation and a discussion on these other functions will ensue. Both charged and uncharged tRNAs have various functions [171]. The uncharged tRNA molecules have been found to play a role in regulating global gene expression. This has been specifically seen in bacteria, in response to amino acid starvation. During amino acid starvation, in yeast and in mammals, uncharged tRNAs activates Gcn2p protein kinase by binding to a specific domain named histidyl-tRNA synthetase (HisRS) domain. This activated domain phosphorylates the translation initiation factor and thereby reduces global protein synthesis [172]. In all these, uncharged tRNAs are merely helping the cells to survive under nutritional stress conditions. Apart from their significance in protein synthesis, they also serve as intermediates in protein degradation mechanisms [173]. tRNAs have also found their way in regulating cell death. They prevent the binding of cytochrome c released from the mitochondria to apoptotic protease activating factor 1 (Apaf1) and thus inhibits the cascade of apoptotic events [174].

One of the recently discovered functions of tRNAs is their processing to generate tRNA fragments (tRFs) [175]. Although these tRFs were initially considered to be degradation by-products, significant roles have been identified for these tRFs, illustrating the fact these fragments are functional molecules that predominantly arise during stress conditions [176]. Their roles have been identified in translational regulation during stress conditions and have also been identified as regulators of gene expression, in a manner similar to that of miRNAs. Relative variations in expression levels of tRFs in tumor cells as compared to normal cells [177] and their role in silencing gene expression, thereby influencing cell proliferation [175] or metastasis [178] implies that they may also contribute to tumorigenesis. Reviews by Keam et al [179] and Shigematsu et al [180] have explained in detail the types of tRFs, their generation and the functional significance. Interestingly, there is also evidence indicating that tRF may possess characteristics of a miRNA, both structurally and functionally (by regulating gene expression) [181], thus expanding the potential repertoire of tRNA functions. Based on the recent discoveries we now understand that tRNAs may also potentially contribute to gene regulation.

1.2.3.4 Clinical relevance of tRNAs

tRNAs have remained as a challenge for biologists as these are not amenable for large scale profiling until recently. However, components of the tRNA biosynthetic machinery such as tRNA synthetases have been found to be deregulated in tumor conditions [182-185]. Structural intricacies and complexities possessed by tRNAs has deterred the development of standard profiling platforms, which in turn has limited our knowledge on the clinical importance of tRNAs. In 2006, Dittmar et al. designed the first microarray chip for tRNAs that captured tRNAs representative of all the amino acids.

Tissue specific expression of tRNAs was also observed in this study [186]. An insight into the contribution of tRNAs to tumorigenesis was gained from the pioneering work by Pavon-Eternod et al [187]. In this study, tRNAs were found to be over expressed in breast tumors and it was suggested that tRNAs may show potential as biomarkers for BC [187]. Following this study, another study was attempted to understand the functional consequences of tRNA overexpression in BC. Overexpression of initiator tRNA ($tRNA_i^{Met}$) was found to promote cell proliferation and increase the metabolic activity of cells [188]. From these studies, it is clear that our knowledge on the clinical importance of tRNAs is primitive. Although tRNAs have been suggested to show promise as biomarkers, no study has been attempted till date to identify tRNAs as prognostic, diagnostic or predictive markers.

1.2.4 Small nucleolar RNAs (snoRNAs)

snoRNAs are one of the most abundant classes of sncRNAs that ranges from 60-300 nt in length [161]. They are a highly conserved group of RNAs that are involved in post-transcriptional modifications and in maturation of other RNAs such as ribosomal RNAs (rRNAs) and small nuclear RNAs (snRNAs). Approximately about 200 different species of snoRNAs are present in every vertebrate cell [189].

1.2.4.1 Genomic organization of snoRNAs

As the name suggests, snoRNAs originate from the nucleolus of a cell, a dynamic organelle that is found within the nucleus of a cell and is involved in rRNA biogenesis and in cell cycle. snoRNAs are mostly encoded within the introns of protein coding and non-protein coding genes or have been found in the intergenic regions [161,190].

More than 90% of human snoRNAs originate in the introns [190]. Usually one intron will harbor one snoRNA. In other organisms such as yeast, metazoans and plants, they may be found as independent genes, as gene clusters or as intronic gene clusters [191]. snoRNAs are most commonly associated with proteins and form small nucleolar ribonucleoprotein complexes [192].

1.2.4.2 snoRNA families

Most of the snoRNAs identified so far, fall into one of the two classes of snoRNA family: C/D box and H/ACA box [191,192]. The former class of snoRNAs are named as SNORDs and the latter are named as SNORAs. SNORDs are involved in 2'O methylation and are associated with four proteins. Fibrillarin is the core protein that is responsible for methylation. C/D box snoRNAs share two sequence motifs which share complementary sequences – C box (PuUGAUGA) and D box (CUGA) at their 5' and 3' ends, respectively. SNORAs are involved in pseudouridylation and are also associated with four proteins. The core protein responsible for this modification is dyskerin. This family of snoRNAs share the sequence motifs H box (ANANNA) and ACA box (ACA) and are characterized by hairpin-hinge-hairpin-tail structure. Small Cajal body RNAs (scaRNAs) are the third class of snoRNAs that are located in the Cajal bodies and are involved in methylation and pseudouridylation of RNA polymerase II transcribed spliceosomal RNAs [193]. All three classes of snoRNAs perform their function by base pairing with target RNAs; while another class of snoRNAs called "orphan snoRNAs" exist that do not share complementarity with any RNA. These post-transcriptional modifications are important to enhance the stability of RNAs and to protect the RNAs from hydrolytic degradation [194].

1.2.4.3 Biogenesis of snoRNAs

snoRNAs existing as independent genes are first transcribed by RNA polymerase II. These transcribed units are excised from both the ends by exonucleases until the snoRNA boundaries are reached. Intronic snoRNAs can mature in two ways – (i) they are excised as lariats during splicing of the host genes. The lariats are debranched and trimmed from both the ends by exonucleases until the snoRNA boundaries are reached. In the second pathway, introns are not excised by splicing but are instead cleaved by endonucleolytic enzymes, which are then trimmed by exonucleases [192].

1.2.4.4 Functions of snoRNAs

The well-known function of snoRNA is in rRNA processing and maturation [191]. Other novel functions of snoRNAs are slowly coming to the fore and some of them are outlined below. Deep sequencing generated data has revealed that processing of snoRNAs may yield other smaller RNAs including miRNAs and piRNAs [161,195-197]. Since miRNAs are considered as master regulators of gene expression, snoRNAs may indirectly be believed to be involved in gene regulation. Recent discoveries have demonstrated the involvement of snoRNAs in alternative splicing [198]. SNORD115 was shown to share complementarity with exon Vb region of serotonin receptor. Vb region contains silencer for splicing, as a result of which this exon is not included in the mRNA, resulting in shorter product. Whereas base pairing of SNORD115 with exon Vb, eliminates the action of the silencer, permitting the inclusion of exon Vb, resulting in a normal receptor. Other mechanisms of how snoRNAs regulate alternative splicing have been proposed but conclusive results are yet to be obtained. One of the indirect mechanisms through which snoRNAs participate in gene regulation is through its

processing to other sncRNAs such as miRNAs and piRNAs. The pathological importance of snoRNAs began to be understood from the observation that a genetic locus containing SNORD115 and SNORD116 was deleted in the neurodevelopmental genetic disorder: the Prader Willi syndrome [199]. Subsequently, other studies have demonstrated altered levels of snoRNAs, indicating their possible involvement in disease conditions, including malignancies. Similar to miRNAs, snoRNAs can also be classified into oncogenes and tumor suppressors, based on their over or under expression in tumor cells, relative to normal. snoRNA deregulation has been observed in metabolic stress disorder [200] and in chronic conditions such as chronic lymphocytic leukemia [201], hepatocellular carcinoma [202], colorectal cancer [203] and endometrial cancer [204], prostate cancer among others. Their roles have also extended to being biomarkers – diagnostic and prognostic. Their diagnostic significance has been observed for lung cancer [205] and their prognostic relevance has been highlighted in colorectal cancer [203], lung cancer [205,206], chronic lymphocytic leukemia [201], peripheral T cell lymphoma [207]. Elevated levels of snoRNA biogenesis has been observed in breast cancer and their significance in breast tumorigenesis has been demonstrated [208,209], yet a comprehensive study on identifying snoRNAs as prognostic markers for BC has not been published so far.

1.3 Profiling platforms for small non-coding RNAs

A number of gene expression profiling platforms have been adopted for small RNAs also. However, small RNAs pose several challenges in developing a profiling platform [210-213] - (i) small size of these RNAs makes it difficult to design a complementary probe or a traditional primer, where often, the size of a probe/primer is

equal to or more than the size of small RNAs, (ii) miRNAs, for instance can differ by a nucleotide and this distinction is difficult to obtain unless the platform is highly sensitive to detect even one nucleotide difference, (iii) The GC content of miRNAs vary greatly, thus making it difficult to standardize the melting temperatures for annealing reactions in a genome wide study, (iv) rapid rate of discovery, making it difficult to reuse the data generated on platforms using pre-printed probes, based on the existing annotation.

Nevertheless, three main platforms used for profiling small non-coding RNAs include microarray, quantitative Reverse Transcription-Polymerase Chain Reaction (qRT-PCR) and next generation sequencing (NGS) [214]. Every platform has its own merits and demerits and a summary of these platforms are provided below.

1.3.1 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

One of the commonly adopted techniques is qRT-PCR that relies on cDNA generated from RNA, followed by real time PCR that quantifies the product in real time. While qRT-PCR offers the advantage of being highly sensitive and specific with a high dynamic range (six orders of magnitude), only limited number of RNAs can be interrogated on this platform [211]. It is expensive, labour intensive and the optimal reaction conditions may vary according to sequence specific differences. The difficulty in designing optimal probes for detecting small RNAs, and the dependence on a pre-determined set of RNAs based on a specific genomic build makes this incompatible for large scale profiling of RNAs. However, this platform may be best suited to validate or probe for candidate molecules and especially when the sample amount in limiting [215].

1.3.2 Microarray

This is a hybridization technique that relies on pairing of RNAs to its complementary sequences printed on a chip. The problem of small size of RNAs poses challenges [212]. Microarrays are less sensitive when compared to qRT-PCR and as with qRT-PCR [210,211], the analysis might be restricted to a single class of molecules, in a specific genome build. A platform with low sensitivity will generate many false negative calls, whereas higher sensitivity and reduced specificity would result in a higher number of false positives. The chance of identifying novel RNAs is minimal and capturing of RNAs with single nucleotide differences is challenging. The dynamic range of this platform is moderate (four orders of magnitude) but it allows profiling of higher number of molecules (compared to qRT-PCR) at a lesser cost [211,216,217]. Replication and comparison of previous study findings generated from using microarray is difficult due to differences in the RNA content printed on the chip. The lack of ability to perform absolute quantification of molecules renders it more suitable for comparing relative abundances of molecules that fall within the dynamic range of the platform between two conditions such as normal and diseased.

Although both these techniques (qRT-PCR and microarray) have their own advantages, these methods rely on a pre-determined set of RNAs based on a specific genome build, thereby leaving us blinded to the functions of other RNAs that cannot be captured because of platform limitation.

1.3.3 Next generation sequencing (NGS)

NGS refers to sequencing of millions of reads in parallel, yielding higher throughput [217] and more coverage, necessitating the use of powerful computing skills and algorithms for analysis [213]. Several platforms are available within NGS. The platform used for the current study is Illumina Genome Analyzer IIx. An overview of Illumina sequencing is as follows [213,218,219]: sequencing in Illumina takes place on a solid glass surface called as flow cell which has sequencing templates (primers). Total RNA is isolated from the sample and is size fractionated and the band corresponding to the size of small RNAs is extracted (~200 bp). Alternatively size selection can also be performed following the addition of adapters. Adapters which are short known sequences of DNA and are complementary to the lawn of primers found on the surface of the flow cell are added to both the 3' and 5' ends of the isolated RNA. Adapters function as primer binding sites for reverse transcription and for PCR amplification. RNA with adapters has to be specifically selected to remove the adapter dimers formed, if any. An agarose gel is again run to select the band corresponding to RNA-adapter to be used for sequencing. The length of RNA to be inserted between the two adapters is user specific and is called the 'insert length'. Sequencing is carried out in a flow cell which has 8 lanes. Each lane is split into two columns and each column is further divided into tiles. There are about 100 tiles per lane. Either one sample can be loaded per lane or multiple samples can be loaded in a single lane where every sample RNA is ligated with an 'index sequence'. Index sequence is very similar to barcode that helps in identifying the respective products. Similarly, index sequence helps in identifying the samples even after they are multiplexed. The process of loading multiple samples in a single lane is called

'multiplexing'. Adapter sequences in the samples hybridize to the primers present on the flow cell. Following hybridization, clusters (clones of the same sequence of RNA) are generated for each and every RNA sequence. This process is called 'cluster generation' and clones are generated by an amplification process called bridge 'amplification'. This process generates about 1000 identical copies of every single RNA template. Clusters are generated to intensify the signal emitted by the fluorescent labeled nucleotides added while sequencing. Sequencing is done using reversible terminator technology, also called as 'sequencing by synthesis'. This technology uses 4 fluorescently labeled nucleotides to sequence millions of clusters in the flow cell surface in parallel. The 3' ends of the nucleotides are reversibly terminated i.e., modified in such a way that the 3'ends are blocked to prevent any further addition of nucleotides. During each sequencing cycle, a single fluorescently labeled nucleotide is added to the growing chain. Soon after the addition of a single nucleotide, the label is imaged to identify the base and the terminator is enzymatically cleaved to allow the incorporation of the next nucleotide.

Based on the insert length and the sequencing direction, two types of sequencing can be done in Illumina – Single end sequencing and Paired end sequencing. Single end sequencing is commonly used for short insert lengths and the sequencing proceeds in only one direction. For small RNA expression profiles, single end sequencing is adopted. Paired end sequencing is done for slightly longer insert lengths (200-600bp) and the sequencing proceeds in both the directions (forward and reverse strands). Mate pair sequencing is done for generating libraries with longer insert lengths (2-5kb).

The advantage of NGS is that it offers absolute quantification of molecules, higher coverage, high sensitivity and specificity [214]. It does not require the knowledge of genomic annotation (prerequisite for qRT-PCR and microarray) and the reads can be assembled de novo [220]. This platform overcomes the problems of hybridization encountered in sequencing technique, is capable of capturing reads with even a single nucleotide difference and is useful for identifying novel RNAs. It exhibits high dynamic range ($\geq$ 10 orders of magnitude), enabling quantification of low amounts of molecules and allows parallel quantification of multiple RNA types, not restricted to one particular class. Since NGS does not depend on any particular genome build, reanalysis of the existing data based on the current genome build is possible. However, sequencing biases may be introduced due to the number of steps involved in sample preparation; data analysis and interpretation is complex due to the large volumes of data being generated. Of the three profiling platforms, NGS is the also the most expensive but it off-sets the costs by allowing mining of all small RNA classes which is not possible on microarray or qRT-PCR platforms. With several user-friendly bioinformatics platforms now available for data analysis, complexity of data and its mining once considered a limitation for NGS has now been overcome [213]. The rationale for my choice of NGS as the profiling platform stems from the above considerations.

1.4 Rationale to conduct the study

sncRNAs are attractive molecules of interest for reasons mentioned below:

Extracting long RNAs from easily available FFPE blocks has been challenging. However, due to their small size, sncRNAs have been demonstrated to be highly stable in

nature, withstanding the effects of formalin and other tissues processing effects. The expressions of these molecules have also remained invariant between fresh frozen and FFPE tissues [221-225]. These properties make them attractive as large repositories of FFPE blocks are housed in pathology departments and when combined with clinical data, are ideal for biomarker discovery and validation studies.

Compared to messenger RNAs, the number of sncRNAs identified so far are fewer in number, which makes the understanding about these molecules and handling of the datasets fairly easy.

sncRNAs are less prone to gene variant mechanisms such as alternative splicing. Therefore analyzing these molecules are less complicated.

sncRNAs have been isolated from almost all the tissue types and biofluids such as serum and plasma and have demonstrated to be highly stable in these biospecimens [205,226-228]. Therefore developing less invasive and easily procurable markers for BC (and other cancers or diseases) prognosis seems plausible.

Lastly, sncRNAs work a step higher in the hierarchy of gene regulation and signaling pathways. The pleiotropic nature and/or redundant properties of sncRNAs draws more attention towards these molecules as altering the expression of a single RNA may have substantial effects on gene expression networks; suggesting that these RNAs may also be helpful for therapeutic interventions.

1.5 Hypothesis

Deregulation of small non-coding RNAs contributes to inter-individual differences in disease trajectory and eventual treatment outcomes in breast cancer.

1.6 Objectives

(i) To comprehensively profile and identify differentially expressed small non-coding RNAs from normal breast tissues and breast tumor tissues.

(ii) To identify miRNAs associated with prognoses (Outcomes: overall survival and recurrence free survival).

(iii) To identify prognostic relevance of sncRNAs profiled (piRNAs, tRNAs and snoRNAs).

Materials and methods common for all the small non-coding RNAs will be explained in chapter 2 and methods specific for each and every RNA will be explained in their respective chapters. The three specific objectives mentioned above have been organized into different chapters. Objective 1 is elaborated in chapter 3 of this thesis and objective 2 is explained in chapter 4. Identification of other small non-coding RNAs i.e., the piRNAs, tRNAs and snoRNAs (objective 3) have been explained in chapters 5, 6 and 7, respectively for more clarity. Further, overall discussion, conclusions and the potential future work are outlined in chapter 8 followed by appendix (chapter 9).

1.7 References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. 2013; Available from: http://globocan.iarc.fr [Cited 2016 Jun].

2. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2015. Toronto, ON: Canadian Cancer Society; 2015. 2015; Available from: http://www.cancer.ca/en/cancer-information/cancer-101/canadian-cancer-statistics-publication/?region=on [Cited 2016 Jun].

3. Weigel MT. Current and emerging biomarkers in breast cancer: Prognosis and prediction. Endocr Relat Cancer 2010;17:R245-262.

4. Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. Lancet 2001;358 North American Edition:1389-1399.

5. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet oncol 2001;2:133-140.

6. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer: Analyses of cohorts of twins from sweden, denmark, and finland. N Engl J Med 2000;343:78-85.

7. McPherson K, Steel CM, Dixon JM. ABC of breast diseases: Breast cancer -- epidemiology, risk factors, and genetics. BMJ 2000;321:624-628.

8. García-Closas M, Brinton LA, Lissowska J, Chatterjee N, Peplonska B, Anderson WF, et al. Established breast cancer risk factors by clinically important tumour characteristics. Br J Cancer 2006;95:123-129.

9. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: Reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94:606-616.

10. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997:1943-1947.

11. Malkin D, Li FP, Strong LC, Fraumeni, Joseph F., Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 1990:1233-1238.

12. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 2006:873-875.

13. American Cancer Society. Breast Cancer Facts & Figures 2015-2016. Atlanta: American Cancer Society, Inc. 2015; Available from: http://www.cancer.org/research/cancerfactsstatistics/breast-cancer-facts-figures [Cited 2016 Jun].

14. Bombardieri E, Gianni L, Bonadonna G, Fabbri A, Carcangiu ML, Carbone A. Histological classification of breast cancer. Breast Cancer - Nuclear Medicine in Diagnosis and Therapeutic Options 2008; p. 3.

15. Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Raymond RS, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: Gene expression analyses across three different platforms. BMC Genomics 2006;7:127.

16. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001(19):10869-10874.

17. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003:8418-8423.

18. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010;28:1684-1691.

19. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.

20. Eroles P, Bosch A, Alejandro Pérez-Fidalgo J, Lluch A. Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. Cancer Treat Rev 2012;38:698-707.

21. Engstrøm,M.J., Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. Breast Cancer Res Treat 2013;140:463-473.

22. Italiano A. Prognostic or predictive? it's time to get back to definitions! J Clin Oncol 2011;29:4718; author reply 4718-4719.

23. Breast cancer facts & figures. [electronic resource]. : Atlanta, Ga. : American Cancer Society [Cited 2016 June].

24. de Boer M, van Dijck JA, Bult P, Borm GF, Tjan-Heijnen VC. Breast cancer prognosis and occult lymph node metastases, isolated tumor cells, and micrometastases. J Natl Cancer Inst 2010;102:410-425.

25. Gospodarowicz MK. Prognostic factors in cancer. : New York : Wiley-Liss, c2001; 2nd ed; 2001.

26. Lønning,P.E. Breast cancer prognostication and prediction: Are we making progress? Ann Oncol 2007;18 Suppl 8:viii3-viii7.

27. Ko A, Dollinger M, Rosenbaum EH. Everyone's guide to cancer therapy : How cancer is diagnosed, treated, and managed day to day. : Kansas City : Andrews McMeel Pub., c2008; Rev. 5th ed; 2008.

28. Moffat FL. Chapter 28: Clinical and pathologic prognostic and predictive factors. in Harris JR, Lippman ME, Morrow M, Osborne CK. Diseases of the Breast, 5th edition, Lippincott Williams & Wilkins, 2014.

29. Gökmen-Polar Y, Badve S. Breast cancer prognostic markers: Where are we now? MLO: Medical Laboratory Observer 2012;44:22-25.

30. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-2826.

31. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care? Nature Reviews Cancer 2007;7:545-553.

32. Haque R, Ahmed SA, Inzhakova G, Shi J, Avila C, Polikoff J, et al. Impact of breast cancer subtypes and treatment on survival: An analysis spanning two decades. Cancer Epidemiology, Biomarkers & Prevention 2012;21:1848-1855.

33. Aparicio S, Mardis E. Tumor heterogeneity: Next-generation sequencing enhances the view from the pathologist's microscope. Genome Biol 2014;15:463-463.

34. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao YJ, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395-399.

35. Ross JS, Hatzis C, Symmans WF, Pusztai L, Hortobágyi G,N. Commercialized multigene predictors of clinical outcome for breast cancer. Oncologist 2008;13:477-493.

36. Ross JS. Multigene predictors in early-stage breast cancer: Moving in or moving out? Expert Review of Molecular Diagnostics 2008:129-135.

37. Ring BZ, Seitz RS, Beck R, Shasteen WJ, Tarr SM, Cheang MC, et al. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3039-3047.

38. O'Hurley G, Sjosted E, Rahman A, Li B, Kampf C, Ponten F, et al. Garbage in, garbage out: A critical evaluation of strategies used for validation of immunohistochemical biomarkers. Molecular oncology 2014;8:783-798.

39. Davis LM, Harris C, Tang L, Doherty P, Hraber P, Sakai Y, et al. Amplification patterns of three genomic regions predict distant recurrence in breast carcinoma. Journal of molecular diagnostics 2007;9:327-336.

40. Sparano JA, Gray RJ, Makower DF, Pritchard K, Albain KS, Hayes DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. N Engl J Med 2015;373:2005-2014.

41. Arranz EE, Vara JÁF, Gámez-Pozo A, Zamora P. Gene signatures in breast cancer: Current and future uses. Translational Oncology 2012;5:398-403.

42. Dabbs DJ, Klein ME, Mohsin SK, Tubbs RR, Shuai Y, Bhargava R. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the oncotype DX test: An independent quality assurance study. J Clin Oncol 2011;29:4279-4285.

43. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, ki-67, and human epidermal growth

factor receptor 2 immunohistochemical score and comparison with the genomic health recurrence score in early breast cancer. J Clin Oncol 2011;29:4273-4278.

44. Parker JS, Mullins M, Cheang M, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology 2009;27:1160-1167.

45. Goetz MP, Suman VJ, Ingle JN, Nibbe AM, Visscher DW, Reynolds CA, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. Clinical cancer research 2006;12:2080-2087.

46. Garber K. Gene expression tests foretell breast cancer's future. Science 2004:1754-1755.

47. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast cancer research 2006;8:R23.

48. van't Veer L, Dai HY, van dV, He Y, Hart A, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-536.

49. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.

50. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nature Reviews Cancer 2004:309-314.

51. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183-1192.

52. Bogaerts J, Cardoso F, Buyse M, Braga S, Loi S, Harrison JA, et al. Gene signature evaluation as a prognostic tool: Challenges in the design of the MINDACT trial. Nature clin pract oncol 2006;3:540-551.

53. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-679.

54. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217-226.

55. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. Women's Oncology Review 2004;4:211-212.

56. Rouzier R, Pusztai L, Delaloge S, Gonzalez-Angulo A, Andre F, Hess KR, et al. Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. J Clin Oncol 2005;23:8331-8339.

57. Damaraju S, Driga A, Cook L, Calder K, Graham K, Dabbs K, et al. 3.02: Considerations on dealing with tissues and cell samples (include tissue banking). Comprehensive Sampling and Sample Preparation 2012:21-31.

58. SEER Stat Fact Sheets: Female Breast Cancer; 2016 [Cited 2016 Jun]. Available from: http://seer.cancer.gov/statfacts/html/breast.html.

59. Abe O, A., Abe R, A., Enomoto K, A., Kikuchi K, A., Koyama H, A., Masuda H, A., et al. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. Lancet 2005:1687-1717.

60. Gonzalez-Angulo A, Morales-Vasquez F, Hortobagyi GN. Overview of resistance to systemic therapy in patients with breast cancer. Adv Exp Med Biol 2007;608:1-22.

61. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 2004;116:843-854.

62. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 2004;116:855-862.

63. Lee RC, Ambros V. An extensive class of small RNAs in caenorhabditis elegans. Science 2001:862-864.

64. Bartel DP. Review: MicroRNAs. genomics, biogenesis, mechanism, and function. Cell 2004;116:281-297.

65. Reinhart BJ, Slack FJ. The 21-nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans. Nature 2000;403:901-906.

66. Griffiths-Jones S. The microRNA registry. Nucleic Acids Res 2004;32:D109-D111.

67. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006;34:D140-D144.

68. Fazi F, Nervi C. MicroRNA: Basic mechanisms and transcriptional regulatory networks for cell fate determination. Cardiovasc Res 2008;79:553-561.

69. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001;294:853-858.

70. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in caenorhabditis elegans. Science 2001;294:858-862.

71. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: Stepwise processing and subcellular localization. EMBO J 2002;21:4663-4670.

72. Zhang X, Azhar G, Wei JY. The expression of microRNA and microRNA clusters in the aging heart. PLoS ONE 2012;7:1-13.

73. Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009;11:228-234.

74. Hausser J, Zavolan M. Identification and consequences of miRNA-target interactions--beyond repression of gene expression. Nature Reviews Genetics 2014;15:599-612.

75. Ramalingam P, Palanichamy JK, Singh A, Das P, Bhagat M, Kassab MA, et al. Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. RNA 2014;20:76-87.

76. Starega-Roslan J, Koscianska E, Kozlowski P, Krzyzosiak WJ. The role of the precursor structure in the biogenesis of microRNA. Cellular and molecular life sciences 2011;68:2859-2871.

77. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III drosha initiates microRNA processing. Nature 2003;425:415-419.

78. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nature Structural & Molecular Biology 2006;13:1097-1101.

79. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and rniRNAs exhibit strand bias. Cell 2003;115:209-216.

80. Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell 2003;115:199-208.

81. Nilsen TW. Mechanisms of microRNA-mediated gene regulation in animal cells. Trends in Genetics 2007;23:243-249.

82. Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. Genes Dev 2004;18:504-511.

83. Berindan-Neagoe I, Monroig PD, Pasculli B, Calin GA. MicroRNAome genome: A treasure for cancer diagnosis and therapy. Ca-A cancer journal for clinicians 2014;64:311-336.

84. Sayed D, Abdellatif M. MicroRNAs in development and disease. Physiol Rev 2011;91:827-887.

85. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res 2005;65:6029-6033.

86. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A 2005;102:13944-13949.

87. Chen C, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science 2004;303:83-86.

88. Jian-Fu Chen, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 2006;38:228-233.

89. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834-838.

90. Volinia S, Calin GA, Liu C, Ambs S, Cimmino A, Petrocca F, et al. A MicroRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103:2257-2261.

91. Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 2005;436:214-220.

92. van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, et al. A signature pattern of stress-responsive MicroRNAs that can evoke cardiac hypertrophy and heart failure. Proc Natl Acad Sci U S A 2006; 103:18255-18260.

93. Liu D, Zhao H, Zhao S, Wang X. MicroRNA expression profiles of peripheral blood mononuclear cells in patients with systemic lupus erythematosus. Acta Histochem 2014; 116:891-897.

94. Stanczyk J, Pedrioli DM, Brentano F, Sanchez-Pernaute O, Kolling C, Gay RE, et al. Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. Arthritis Rheum 2008;58:1001-1009.

95. Zovoilis A, Agbemenyah HY, Agis-Balboa R, Stilling RM, Edbauer D, Rao P, et al. microRNA-34c is a novel target to treat dementias. EMBO J 2011;30:4299-4308.

96. Kim J, Inoue K, Ishii J, Vanti WB, Voronov SV, Murchison E, et al. A MicroRNA feedback circuit in midbrain dopamine neurons. Science 2007;317:1220-1224.

97. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 2002;99:15524-15529.

98. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res 2004;64:3753-3756.

99. Venkataraman S, Birks DK, Balakrishnan I, Alimova I, Harris PS, Patel PR, et al. MicroRNA 218 acts as a tumor suppressor by targeting multiple cancer phenotype-associated genes in medulloblastoma. J Biol Chem 2013;288:1918-1928.

100. Li XJ, Ren ZJ, Tang JH. MicroRNA-34a: A potential therapeutic target in human cancer. Cell death & disease 2014;5:e1327.

101. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065-70.

102. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res 2004;64:3087-3095.

103. Papagiannakopoulos T, Shapiro A, Kosik KS. MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. Cancer Res 2008;68:8164-8172.

104. Voorhoeve PM, le Sage C, Schrier M, Gillis AJM, Stoop H, Nagel R, et al. Article: A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell 2006;124:1169-1181.

105. Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al. MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol 2008;26:462-469.

106. Janssen HLA, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K, et al. Treatment of HCV infection by targeting microRNA. N Engl J Med 2013;368:1685-1694.

107. Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-microRNA signature. Gut 2010;59:579-585.

108. Liu N, Cui R, Sun Y, Guo R, Mao Y, Tang L, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. International Journal of Cancer 2014;134:1359-68.

109. Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

110. Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics 2015;16:735-735.

111. Shin VY, Ng EKO, Chan VW, Kwong A, Kent-Man Chu. A three-miRNA signature as promising non-invasive diagnostic marker for gastric cancer. Molecular Cancer 2015;14:1-9.

112. Luo J, Chen M, Huang H, Yuan T, Zhang M, Zhang K, et al. Circulating microRNA-122a as a diagnostic marker for hepatocellular carcinoma. OncoTargets & Therapy 2013;6:577-583.

113. Motawi TMK, Rizk SM, Shaker OG, Mokhtar OZH. MicroRNAs as predictor markers for response to interferon treatment of chronic hepatitis C genotype-4 in egyptian patients. PLoS ONE 2015;10:1-12.

114. Lopes-Ramos C, Habr-Gama A, Quevedo BD, Felcio NM, Bettoni F, Koyama FC, et al. Overexpression of miR-21-5p as a predictive marker for complete tumor regression to neoadjuvant chemoradiotherapy in rectal cancer patients. Bmc medical genomics 2014;7:68.

115. Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-1184.

116. Pérez-Rivas LG, Jerez JM, Carmona R, de Luque V, Vicioso L, Claros MG, et al. A microRNA signature associated with early recurrence in breast cancer. PLoS ONE 2014;9:1-14.

117. Cuk K, Zucknick M, Heil Jö, Madhavan D, Schott S, Turchinovich A, et al. Circulating microRNAs in plasma as early detection markers for breast cancer. Int J Cancer 2013;132:1602-1612.

118. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 2013;19:4477-4487.

119. Suzuki R, Honda S, Kirino Y. PIWI expression and function in cancer. Front Genet 2012;3:204-204.

120. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian piwi proteins. Nature 2006;442:199-202.

121. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 2006;442:203-207.

122. Grivna ST, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20:1709-14.

123. Watanabe T, Lin H. Posttranscriptional regulation of gene expression by piwi proteins and piRNAs. Mol Cell 2014;56:18-27.

124. Deng W, Lin HF. Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Developmental cell 2002;2:819-830.

125. Kuramochi-Miyagawa S, Kimura T, Yomogida K, Kuroiwa A, Tadokoro Y, Fujita Y, et al. Two mouse piwi-related genes: Miwi and mili. Mech Dev 2001;108:121-133.

126. Kim VN. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. Genes Dev 2006;20:1993-1997.

127. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in drosophila. Cell 2007;128:1089-1103.

128. Le Thomas A, Tóth KF, Aravin AA. To be or not to be a piRNA: Genomic origin and processing of piRNAs. Genome Biol 2014;15:204.

129. Martinez VD, Vucic EA, Thu KL, Hubaux R, Enfield KSS, Pikor LA, et al. Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. Sci Rep 2015;5:10423-10423.

130. Luteijn MJ, Ketting RF. PIWI-interacting RNAs: From generation to transgenerational epigenetics. Nature Reviews Genetics 2013;14:523-534.

131. Assumpcao CB, Calcagno DQ, Araujo T, dos Santos S, dos Santos A, Riggins GJ, et al. The role of piRNA and its potential clinical implications in cancer. Epigenomics 2015;7:975-984.

132. Thomson T, Lin HF. The biogenesis and function of PIWI proteins and piRNAs: Progress and prospect. Annu Rev Cell Dev Biol 2009;25:355-376.

133. Mei YP, Clark D, Mao L. Novel dimensions of piRNAs in cancer. Cancer Lett 2013;336:46-52.

134. Zamore PD. Somatic piRNA biogenesis. EMBO J 2010;29:3219-3221.

135. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: The vanguard of genome defence. Nature Reviews Molecular Cell Biology 2011;12:246-258.

136. Rouget C, Papin C, Boureux A, Meunier A, Franco B, Robine N, et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early drosophila embryo. Nature 2010;467:1128-1132.

137. Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Alvarado AS. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science 2005;310:1327-1330.

138. Palakodeti D, Smielewska M, Lu YC, Yeo GW, Graveley BR. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. RNA 2008;14:1174-1186.

139. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell 2002;110:689-699.

140. Qiao D, Zeeman AM, Deng W, Looijenga LH, Lin HF. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. Oncogene 2002;21:3988-3999.

141. Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA, et al. Heterochromatic silencing and HP1 localization in drosophila are dependent on the RNAi machinery. Science 2004;303:669-672.

142. Yin H, Lin H. An epigenetic activation role of piwi and a piwi-associated piRNA in drosophila melanogaster. Nature 2007;450:304-308.

143. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 2008;31:785-799.

144. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science 2006;313:363-367.

145. Catalanotto C, Azzalin G, Macino G, Cogoni C. Involvement of small RNAs and role of the qde genes in the gene silencing pathway in neurospora. Genes Dev 2002;16:790-795.

146. Girard A, Hannon GJ. Conserved themes in small-RNA-mediated transposon control. Trends Cell Biol 2008;18:136-148.

147. Esposito T, Magliocca S, Formicola D, Gianfrancesco F. piR_015520 belongs to piwi-associated RNAs regulates expression of the human melatonin receptor 1A gene. PLOS ONE 2011;6:e22727.

148. Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-9910.

149. Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, et al. Identification of novel piRNAs in bladder cancer. Cancer Lett 2015;356:561-567.

150. Lee JH, Schütte D, Wulf G, Füzesi L, Radzun H, Schweyer S, et al. Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/bcl-XL pathway. Hum Mol Genet 2006;15:201-211.

151. Lee JH, Jung C, Javadian-Elyaderani P, Schweyer S, Schütte D, Shoukier M, et al. Pathways of proliferation and antiapoptosis driven in breast cancer stem cells by stem cell protein piwil2. Cancer Res 2010;70:4569-4579.

152. Liu JJ, Shen RL, Chen L, Ye Y, He G, Hua KD, et al. Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. International journal of clinical and experimental pathology 2010;3:328-337.

153. Wen-Kang Liu, Xiang-Yang Jiang, Zhen-Xi Zhang. Expression of PSCA, PIWIL1 and TBX2 and its correlation with HPV16 infection in formalin-fixed, paraffin-embedded cervical squamous cell carcinoma specimens. Arch Virol 2010;155:657-663.

154. He G, Chen L, Ye Y, Xiao Y, Hua KD, Jarjoura D, et al. Piwil2 expressed in various stages of cervical neoplasia is a potential complementary marker for p16(INK4a). American journal of translational research 2010;2:156-169.

155. Sun GA, Wang YY, Sun LH, Luo H, Liu N, Fu Z, et al. Clinical significance of hiwi gene expression in gliomas. Brain Res 2011;1373:183-188.

156. Grochola LF, Greither T, Taubert H, Moller P, Knippschild U, Udelnow A, et al. The stem cell-associated hiwi gene in human adenocarcinoma of the pancreas: Expression and risk of tumour-related death. Br J Cancer 2008;99:1083-1088.

157. Zeng Y, Qu LK, Meng L, Liu CY, Dong B, Xing XF, et al. HIWI expression profile in cancer cells and its prognostic value for patients with colorectal cancer. Chin Med J 2011;124:2144-2149.

158. Huang G, Hu H, Xue X, Shen S, Gao E, Guo G, et al. Altered expression of piRNAs and their relation with clinicopathologic features of breast cancer. Clinical & translational oncology 2013;15:563-568.

159. Martinez VD, Enfield KSS, Rowbotham DA, Lam WL. An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence. Gastric Cancer 2016;19:660-665.

160. Busch J, Ralla B, Jung M, Wotschofsky Z, Trujillo-Arribas E, Schwabe P, et al. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. Journal of Experimental & Clinical Cancer Research (17569966) 2015;34:1-11.

161. Martens-Uzunova E, Olvedy M, Jenster G. Beyond microRNA - novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett 2013;340:201-211.

162. Chan PP, Lowe TM. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 2009;37:D93-D97.

163. Crick FH. Codon--anticodon pairing: The wobble hypothesis. J Mol Biol 1966;19:548-555.

164. RajBhandary U, So¨ll D. tRNA : Structure, biosynthesis, and function. : Washington, D.C. : ASM Press, c1995; 1995.

165. Talbert PB, Henikoff S. Chromatin-based transcriptional punctuation. Genes Dev 2009;23:1037-1041.

166. Bermudez-Santana C, Attolini C, Kirsten T, Engelhardt J, Prohaska SJ, Steigele S, et al. Genomic organization of eukaryotic tRNAs. BMC Genomics 2010;11:270.

167. Hoagland MB, keller EB, zamecnik PC. Enzymatic carboxyl activation of amino acids. J Biol Chem 1956;218:345-358.

168. Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC. A soluble ribonucleic acid intermediate in protein synthesis. J Biol Chem 1958;231:241-257.

169. Spears JL, Rubio MAT, Sample PJ, Alfonzo JD. tRNA biogenesis and processing. RNA Metabolism in Trypanosomes 2012:99.

170. Hopper AK, Phizicky EM. tRNA transfers to the limelight. Genes Dev 2003;17:162-180.

171. Raina M, Ibba M. tRNAs as regulators of biological processes. Front Genet 2014;5:171-171.

172. Wek SA, Zhu S, Wek RC. The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. Mol Cell Biol 1995;15:4497-4506.

173. Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. Science 1986;234:179-186.

174. Mei YD, Yong J, Liu HT, Shi YG, Meinkoth J, Dreyfuss G, et al. tRNA binds to cytochrome c and inhibits caspase activation. Mol Cell 2010;37:668-678.

175. Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 2009;23:2639-2649.

176. Wang QR, Lee I, Ren JP, Ajay SS, Lee YS, Bao XY. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. Molecular therapy 2013;21:368-379.

177. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep sequencing of serum small RNAs identifies patterns of 5' tRNA half and YRNA fragment expression associated with breast cancer. Biomark Cancer 2014;6:37-47.

178. Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. Cell 2015;161:790-802.

179. Keam SP, Hutvagner G. tRNA-derived fragments (tRFs): Emerging new roles for an ancient RNA in the regulation of gene expression. Life (2075-1729) 2015;5:1638-1651.

180. Shigematsu M, Honda S, Kirino Y. Transfer RNA as a source of small functional RNA. J Mol Biol Mol Imaging 2014;1.

181. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 2013;110:1404-1409.

182. Bjornsti MA, Houghton PJ. Lost in translation: Dysregulation of cap-dependent translation and cancer. Cancer cell 2004;5:519-523.

183. Chen W, Böcker W, Brosius J, Tiedge H. Expression of neural BC200 RNA in human tumours. J Pathol 1997;183:345-351.

184. Marshall L, White RJ. Non-coding RNA production by RNA polymerase III is implicated in cancer. Nature Reviews Cancer 2008;8:911-914.

185. Winter AG, Sourvinos G, Allison SJ, Tosh K, Scott PH, Spandidos DA, et al. RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. Proc Natl Acad Sci U S A 2000;97:12619-12624.

186. Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. PLoS Genetics 2006;3:2107-2115.

187. Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res 2009;37:7268-7280.

188. Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. RNA 2013;19:461-466.

189. Smith CM, Steitz JA. Sno storm in the nucleolus: New roles for myriad small RNPs. Cell 1997;89:669-672.

190. Dieci G, Preti M, Montanini B. Minireview: Eukaryotic snoRNAs: A paradigm for gene expression flexibility. Genomics 2009;94:83-88.

191. Dragon F, Filipowicz W. SnoRNAs: Biogenesis, structure and function. Encyclopedia of Life Sciences 2005:317.

192. Filipowicz W, Pogacic V. Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol 2002;14:319-327.

193. Gall JG. The centennial of the cajal body. Nature Reviews Molecular Cell Biology 2003;4:975-980.

194. Bachellerie JP, Cavaille J, Huttenhofer A. The expanding snoRNA world. Biochimie 2002;84:775-790.

195. Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, et al. Article: A human snoRNA with MicroRNA-like functions. Mol Cell 2008;32:519-528.

196. Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res 2011;39:675-686.

197. Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, et al. A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. Nucleic Acids Res 2015;43:10474-10491.

198. Kishore S, Stamm S. Regulation of alternative splicing by snoRNAs. Cold Spring Harb Symp Quant Biol 2006;71:329-334.

199. Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, et al. Prader-willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat Genet 2008;40:719-721.

200. Michel CI, Holley CL, Scruggs BS, Sidhu R, Brookheart RT, Listenberger LL, et al. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. Cell metabolism 2011;14:33-44.

201. Ronchetti D, Mosca L, Cutrona G, Tuana G, Gentile M, Fabris S, et al. Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia. BMC Medical Genomics 2013;6:1-11.

202. Xu G, Yang F, Ding CL, Zhao LJ, Ren H, Zhao P, et al. Small nucleolar RNA 113-1 suppresses tumorigenesis in hepatocellular carcinoma. Molecular cancer 2014;13:216.

203. Okugawa Y, Toiyama Y, Toden S, Mitoma H, Nagasaka T, Tanaka K, et al. Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. Gut 2015.

204. Ravo M, Cordella A, Rinaldi A, Bruno G, Alexandrova E, Saggese P, et al. Small non-coding RNA deregulation in endometrial carcinogenesis. Oncotarget 2015;6:4677-4691.

205. Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkersfor non-small-cell lung cancer. Molecular Cancer 2010;9:198-207.

206. Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, et al. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. Int J Cancer 2015;136:E623-E629.

207. Valleron W, Ysebaert L, Berquet L, Fataccioli V, Quelen C, Martin A, et al. Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. Blood 2012;120:3997-4005.

208. Dong X, Guo P, Boyd J, Sun X, Li Q, Zhou W, et al. Implication of snoRNA U50 in human breast cancer. Journal of Genetics and Genomics 2009;36:447-454.

209. Su H, Xu T, Ganapathy S, Shadfan M, Long M, Huang TH, et al. Elevated snoRNA biogenesis is essential in breast cancer. Oncogene 2014;33:1348-1358.

210. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: Approaches and considerations. Nature Reviews Genetics 2012;13:358-369.

211. Baker M. MicroRNA profiling: Separating signal from noise: Various platforms for measuring microRNAs can provide different answers. Nature Methods 2010;7:687-692.

212. Wark AW, Lee HJ, Corn RM. Multiplexed detection methods for profiling MicroRNA expression in biological samples. Angewandte Chemie 2008;47:647-652.

213. Pimentel F, Bonilla P, Ravishankar YG, Contag A, Gopal N, LaCour S, et al. Technology in MicroRNA profiling: Circulating MicroRNAs as noninvasive cancer biomarkers in breast cancer. J Lab Autom 2015;20:574-588.

214. Tam S, de Borja R, Tsao M, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest 2014;94:350-358.

215. Moldovan L, Batte KE, Trgovcich J, Wisler J, Marsh CB, Piper M. Methodological challenges in utilizing miRNAs as circulating biomarkers. J Cell Mol Med 2014;18:371-390.

216. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief Funct Genomic Proteomic 2009;8:174-183.

217. De Cecco L, Dugo M, Canevari S, Daidone MG, Callari M. Measuring microRNA expression levels in oncology: From samples to data analysis. Crit Rev Oncog 2013;18:273-287.

218. Gogol-Döring A, Chen W. An overview of the analysis of next generation sequencing data. Methods Mol Biol 2012;802:249-257.

219. Myllykangas S, Buenrostro J, Ji HP. Overview of sequencing technology platforms. Bioinformatics for High Throughput Sequencing 2012:11.

220. Grada A, Weinbrecht K. Next-generation sequencing: Methodology and application. J Invest Dermatol 2013;133:E1-E4.

221. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. RNA 2007;13:1668-1674.

222. Zhang X, Chen J, Radcliffe T, Lebrun DP, Tron VA, Feilotter H. An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. J Mol Diagn 2008;10:513-519.

223. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol 2007;7:36-36.

224. Bovell L, Shanmugam C, Katkoori VR, Zhang B, Vogtmann E, Grizzle WE, et al. microRNAs are stable in formalin-fixed paraffin-embedded archival tissue specimens of colorectal cancers stored for up to 28 years. Frontiers in bioscience (Elite edition) 2012;4:1937-1940.

225. Peiró-Chova L, Peña-Chilet M, López-Guerrero JA, García-Giménez JL, Alonso-Yuste E, Burgues O, et al. High stability of microRNAs in tissue samples of compromised quality. Virchows archiv 2013;463:765-774.

226. Creemers EE, Tijsen AJ, Pinto YM. Circulating MicroRNAs novel biomarkers and extracellular communicators in cardiovascular disease? Circ Res 2012;110:483-495.

227. Yang X, Cheng YD, Lu Q, Wei JF, Yang HW, Gu M. Detection of stably expressed piRNAs in human blood. International journal of clinical and experimental medicine 2015;8(8):13353-13358.

228. Bahn JH, Zhang Q, Li F, Chan TM, Lin X, Kim Y, et al. The landscape of MicroRNA, piwi-interacting RNA, and circular RNA in human saliva. Clin Chem 2015;61:221-230.

2. Materials and methods

US National Cancer Institute and the European Organisation for Research and Treatment of Cancer (NCI-EORTC) at their First International Meeting on Cancer Diagnostics (From Discovery to Clinical Practice: Diagnostic Innovation, Implementation, and Evaluation), recommended specific guidelines for tumor based biomarker discovery, validation and reporting. McShane et al, laid guidelines on reporting practices, called "REporting recommendations for tumour MARKer prognostic studies" which is well-known as REMARK guidelines [1]. I have adhered to these guidelines in identifying and reporting prognostic markers for BC in this thesis. Briefly, REMARK guidelines call for: (i) explanations about the markers under investigation, a clear statement of the study objectives and hypothesis; (ii) description of the patient cohorts as well as characteristics of the specimens used for the study; (iii) estimation of the sample size needed to conduct the study; (iv) methods adopted, (v) overall study design and the various statistical methods adopted (including the method to estimate optimal cut-off point) and the results thus obtained (which includes hazards ratio, confidence interval); (vi) validation of the initial findings from discovery cohort in an external dataset and (vii) discuss potential limitations and implications of the study for future research.

2.1 Sample size calculation

Number of samples needed to detect statistically significant differences of the measured sncRNAs between the two comparison groups i.e., cases and controls was estimated using the following web tools:

http://bioinformatics.mdanderson.org/MicroarraySampleSize/ and

http://linus.nci.nih.gov/brb/samplesize/ [2,3]. I considered the following parameters to estimate the sample sizes: α (acceptable number of false positives) = 0.05 (5%), β (desired power to conduct the study) = 80% and a fold difference of 2 or more in sncRNA expression. Under these conditions, at least 8-11 samples were required in each group (controls and cases). This study included 11 control samples and 104 cases, thus meeting the statistical requirements to enable data interpretations with confidence.

2.2 Clinical characteristics of the samples used for the study

2.2.1 Discovery cohort

Samples included in the discovery cohort were obtained from women in Alberta, Canada. Written informed consent was obtained from all the study participants and the study was approved by the local Institutional Research Ethics committee (Health Research Ethics board of Alberta- Cancer Committee).

Controls:

Eleven apparently healthy breast tissues obtained from reduction mammoplasty surgery were considered as normal samples and will henceforth be called as controls. These samples were stored as flash frozen (FF) tissue specimens. The tissue samples were assessed by a pathologist and were confirmed to be free of malignancy.

Cases:

Breast tumor tissues from one hundred and four cancer patients diagnosed with invasive ductal breast cancer were obtained to conduct this study and will henceforth be

called as cases. These samples, along with their complete clinical characteristics were accessed from Alberta Cancer Research Biobank/Canadian Breast Cancer Foundation tumor bank (http://www.acrb.ca/). All the patients were non-metastatic at the time of diagnosis (except one) and the samples were collected between the years 1996 and 2008. The median follow up period was 2927.5 days or 8.02 years (range: 170 – 6125 days) and the median age at the time of diagnosis was 50 years (range: 27 – 79 years). Patients were classified into different molecular subtypes based on their immunohistochemical profiles that considered estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor (HER2) expression levels. More than 50% (n = 62) of the patients belonged to Luminal A subtype with positive expression status for ER, PR and negative status for HER2 receptor. Thirty patients were diagnosed with triple negative breast cancer (TNBC) based on low to negligible expression of all three receptors. Ten patients showed positive expression status for all the three receptors and were classified as Luminal HER2 subtype. Two patients were positive for ER and PR and their HER2 status was unknown. Since the overall tumor grade was high, these were classified into Luminal B subtype, as described earlier [4-6]. Luminal HER2 and Luminal B samples will henceforth be called as Luminal B. Of the 104 patients, 25 underwent neoadjuvant therapy and 79 underwent adjuvant therapy, with predominant treatment option being the administration of TAC (Taxotere/docetaxel, Adriamycin and Cyclophosphamide; n = 57). Despite standard care of treatment, 46 patients died and 61 patients experienced recurrence. All the tissue samples were preserved as Formalin Fixed Paraffin Embedded tissue blocks (FFPE). A pathologist (Dr. Richard Berendt) examined tumor cellularity in H&E stained sections from each of these blocks and found that all of the 104 samples

exhibited ≥ 70% tumor cells. The percent distribution of tumor cellularity is as follows: 70% (n = 7), 80 – 90% (n = 13), 90% (n = 24), 95% (n = 35) and 100% (n = 25).

2.2.2 External validation cohort (TCGA)

Samples profiled by The Cancer Genome Atlas (TCGA) and preserved as FF tissues were accessed for use as external validation cohort. The data access committee from TCGA approved the study protocol and use of data sets. A total of 1,088 BC cases were available in TCGA dataset. I filtered the samples based on the following criteria, to make the dataset comparable to the discovery cohort from Alberta (n=104): (i) female patients, (ii) absence of any previous malignancy, (iii) non-metastatic at the time of presentation, (iv) non Caucasian samples were removed based on the self-declared ethnicity, and (v) invasive ductal carcinomas. 479 samples were retained after filtering for the above mentioned criteria. Data on the hormone receptor status was available for 332 patients, using which, I classified the samples into Luminal A (n = 203), Luminal B (n = 58), TNBC (n = 52) and HER2+ enriched (n = 19) [5]. Tumor stage information was available only for 328 samples. Samples in TCGA dataset was sequenced using two platforms – Illumina Genome Analyzer IIx and Illumina HiSeq. Of the 328 samples, 156 were sequenced using the former and 172 were sequenced using the latter. Samples sequenced from Illumina Genome Analyzer would have been the ideal external dataset for comparison with the discovery cohort but the number of events (n = 8) from this subset of samples was less to run a survival analysis, therefore samples sequenced using Illumina HiSeq platform was used as an external dataset. Since the discovery cohort did not include any HER2+ enriched samples, I removed these samples from TCGA dataset, resulting in a sample size of 162. I considered only samples with a follow-up period of >

3 years for patients without any events, based on a previous study that defined the follow-up period for recurrence or survival analysis [5]. Overall, I was left with 84 samples for survival analysis, with death reported for 27 patients. TCGA data lacked time to recurrence and hence I attempted only the survival analysis. The percent distribution of tumor cells (cellularity) in TCGA dataset were as follows: (i) 30 – 50% = 14, (ii) 55 – 70% = 19 and (iii) 75 – 100% = 50. One sample did not have any information on tumor cellularity. Compared to the discovery cohort, in which all the samples had tumor cellularity > 70%, the number of samples with > 70% tumor cellularity were less in TCGA dataset (at 60% of the total n=84). All these differences, including the platform differences were taken into account for finer interpretations of the data.

Patient demographics of discovery and external validation cohorts are summarized in Table 2.1.

Table 2.1. Patient demographics of discovery and external validation cohorts

Characteristics	Discovery cohort from Alberta (n=104)	External validation cohort from TCGA (n=84)
Median age at diagnosis in years (range)	50 (24 – 79)	54.5 (35 – 90)
Median follow up time from diagnosis in days (range)	2927.5 (170 – 6125)	1881.5 (174 – 3807)
Molecular subtypes Luminal A Luminal B Triple Negative	 62 12 30	 51 18 15
Menopausal status Pre Post Peri Unknown	 37 75 11 1	 24 46 3 11
Family history of Breast Cancer Yes No Unknown	 40 58 6	 N/A N/A N/A
Stage I II III IV	 8 79 16 1	 25 47 12 0
Overall Grade Low High Unknown	 36 67 1	 N/A N/A N/A
Vital Status Alive Dead	 58 46	 57 27
Relapse Status Relapse No relapse	 61 43	 N/A N/A
Treatment type Adjuvant Neoadjuvant	 79 25	 84 0

2.3 Isolation of total RNA for small RNA sequencing

I homogenized all the control samples, stored as FF tissues using TRIzol (Invitrogen) and isolated total RNA using Qiagen RNeasy kit according to manufacturers' instructions. Total RNA from FFPE tissues was isolated using RecoverAll Total Nucleic Acid Isolation Kit (Life technologies) through the services of PlantBiosis Ltd (Lethbridge, Alberta, Canada; http://www.plantbiosis.com/). RNA quality and quantity were analyzed with Bioanalyzer 2100 and RNA Nano Chips (Agilent Technologies). The RNA extraction protocols that have been followed for this study have previously been optimized for FF and FFPE tissues wherein the use of different extraction protocols in a comparative miRNA study was shown to result in expression profiles that are highly reproducible and strongly correlated between FF and FFPE tissue types [7,8].

2.4 Small RNA sequencing

Services from PlantBiosis Ltd were utilized preparing small RNA libraries and for small RNA sequencing. Basic bioinformatics support, i.e., from generating fastq files to .bam files were also offered by PlantBiosis Ltd and the details are as follows: Small RNAs were sequenced using TruSeq Small RNA Sequencing Kit (Illumina), TruSeq SR Cluster Kit v5-CS-GA (Illumina) and TruSeq SBS Kit v5-GA (Illumina) according to manufacturer's instructions. This sequencing protocol aims to select and amplify small RNAs, ranging between 15 and 40 nt in length. Therefore, size fractionation was performed to include only sequences less than 200nt, after adapter ligation. All the samples were sequenced on Illumina Genome Analyzer IIx with 36-cycle single-end protocol (7 belonged to index and 29 base sequence for alignment with genome build).

Base calling and demultiplexing were completed using CASAVA 1.8.2 with default settings, followed by trimming of adapters using CutAdapt software (http://code.google.com/p/cutadapt/). Sequences longer than 17 nucleotides and ≤ 27 nucleotides were retained. Quality trimming was performed to retain only reads with a Sanger quality score cut-off of 30. The quality of the sequenced reads after adapter trimming was assessed using FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). One tumor sample was not processed further due to poor quality and was therefore excluded, leaving 103 tumor samples for further analysis. Trimmed sequences were then aligned to the reference genome using Bowtie [9] and were allowed a maximum of two mismatches. Human hg19 genomic assembly (UCSC), downloaded from Illumina iGenome repository was used as a reference for mapping. Aligned sequences were saved as .sam files, converted to more memory efficient .bam files and sorted by genomic position. Sequencing data was submitted to Gene Expression Omnibus (GEO accession ID GSE68085).

2.5 Sequencing data analysis

2.5.1 Discovery cohort

For analysis of NGS data, I used Partek Genomics Suite v 6.6 (PGS, Partek® Genomics Suite software, Version 6.6 beta, Copyright © 2009 Partek Inc., St. Louis, MO, USA). The .bam files of 103 tumor samples and 11 normal samples served as input files to PGS. Four classes of sncRNAs, i.e., miRNAs, piRNAs, tRNAs and snoRNAs were studied in this work and individual data analysis was performed for every class of

sncRNA. The following databases were used to annotate the .bam files to different sncRNA classes:

miRBase v20 (http://www.mirbase.org/) for mature miRNAs [10,11],

piRNA bank (http://pirnabank.ibab.ac.in/index.shtml) [12] for piRNAs,

UCSC (http://gtrnadb.ucsc.edu) [13] for tRNAs,

and Ensembl (http://grch37.ensembl.org/index.html) for snoRNAs [14].

For every class of sncRNA, all the sncRNAs that registered a read count of one was annotated from the dataset. The dataset was normalized using reads per kilobase per million method (RPKM), a well-established method used for normalizing sequencing data [15].

Separate small RNA libraries were constructed for different batches of tissue specimens profiled. Hence the datasets were adjusted for potential batch effects using ANOVA model. Overall, the samples used for this study were sequenced in four different batches: Batch 1 = 8 TNBC samples; Batch 2 = 16 Luminal A samples; Batch 3 = 11 normal samples; 10 TNBC samples & 23 Luminal A samples; Batch 4 = 25 Luminal A samples, 10 Luminal B samples & 11 TNBC samples. Sequencing was performed in different batches for the following reasons: (i) to initially explore the feasibility of profiling sncRNA, (ii) as a quality control (QC) step to determine the quality and amenability of specimens available for high-throughput sequencing and (iii) to contain costs at the exploratory stage. Many of the technologies used in biology are often encountered with variations arising from technical and biological factors. Our whole objective of performing an experiment is to capture signals arising from biological source and not from technical source. One of the main sources of variations is called as batch

effects, which, as defined by Leek et al., are "sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study" [16]. One example of a reason for batch effects is the processing of samples on different days. The presence of batch effects in a dataset may lead to inaccurate biological conclusions, creating difficulty in reproducing the results. However, if a dataset that is confounded by batch effects is analyzed, it becomes difficult to distinguish the results thus obtained from those that would arise from real biological effects. Therefore, it is imperative to understand the dataset or in other words, perform an exploratory analysis well before performing the real experiment and interpreting the results. There are several methods to quantify and correct for batch effects and this has explained in detail by Leek et al [16]. One of the methods that I have adopted is to cluster the samples using Principal component analysis (PCA) and quantify the amount of variation arising from the presence of batch effects using ANOVA model. The presence of batch effects warranted data correction for batch effects. The option for correcting for batch effects is in built within PGS.

For subsequent analysis, sncRNAs were filtered for read counts – sncRNAs that had $\geq$ 10 read counts in at least 90% of the samples (normal and tumor samples inclusive for case-control approach and only tumor samples for case-only approach; the two statistical approaches are explained in section 2.6) were retained for all the downstream analysis. PCA plot of filtered raw counts of each class of sncRNA corrected for batch effects was used for identifying potential sample outliers. Samples deviating from three standard deviations were identified as potential sample outliers. After removing sample outliers, .bam files were reloaded into PGS and the dataset was normalized and corrected

for batch effects. Further, the same filtering cut off was applied to retain sncRNAs for downstream analysis. For all the downstream analysis, batch effects corrected normalized counts of filtered sncRNAs was used. One-way ANOVA test was used to identify differentially expressed (DE) sncRNAs with fold change (FC) $\geq$ 2.0 and false discovery rate (FDR) cut off $\leq$ 0.05.

2.5.2 External validation cohort

Eighty four samples from TCGA dataset were considered as the external validation cohort. I analyzed the .bam files of 84 samples using PGS. All the sncRNAs which registered a read count of one were annotated and the dataset was normalized using RPKM method. Further, the dataset was adjusted for batch effects using ANOVA model, considering the following variables: batch ID, plate ID and tissue source site.

2.6 Survival analysis

I performed the statistical analyses under the able guidance of Dr. Sunita Ghosh.

2.6.1 Discovery cohort

Two commonly used statistical approaches were adopted for this study: Case-control (CC) approach and Case-only (CO) approach. The difference between these two approaches lies in the selection procedure of sncRNAs for survival analysis. In the CC approach, only DE sncRNAs were tested for their association with outcomes (Overall Survival, OS and Recurrence Free Survival, RFS). In contrast, CO method is unbiased, i.e., it includes all of the sncRNAs profiled in the tumor samples (following the data filtering criteria described above), thus allowing a wider dataset for interrogation and is not influenced by expression differences between normal and tumor samples, thus

eliminates the bias introduced by the defined “normal sample”. CO approach therefore offers a chance to identify molecules, which would have otherwise been missed in a DE list identified from CC approach and also includes those RNAs which may be preferentially expressed in tumor samples alone and not in normal samples. Here again, all the sncRNAs retained after applying the filter cut off ($\geq$ 10 read counts in at least 90% of the tumor samples) were subjected to survival analysis.

More often than not, high-throughput techniques suffer from the problems of high dimensionality (a higher number of markers but lower number of samples) and collinearity (correlation between two markers), leading to the generation of instable co-efficients in a traditional Cox-proportional hazards regression model [17]. In such cases, the inclusion of individual miRNAs to build a model may not yield reliable results, whereas considering miRNAs as continuous variables and constructing risk scores overcomes both these problems. For both OS and RFS, sncRNAs obtained from CC and CO approaches were considered as continuous variables and were subjected to Univariate cox proportional hazards regression analysis, along with permutation test (n = 10,000) using ‘glmperm’ package in R statistical program. OS and RFS were defined as the time period between the date of surgery and until an event occurred – death in case of OS and recurrence in case of RFS. Any sncRNA that was significant in the permutation test with $p \leq 0.1$ were considered for further analysis. All the subsequent analysis was performed using SAS version 3.3 (SAS institute Inc., Cary, NC). sncRNAs that were significant in the permutation test were used for constructing a risk score for each sample. The formula for risk score construction is as follows:

$Risk\ Score_i = \sum_{j=1}^{12} \beta_j * sncRNA_{ij}$, where $sncRNA_{ij}$ is the individual risk score for sncRNA j on sample i, and β_j is the parameter estimate obtained from the univariate analysis for sncRNA j [18]. For all the sncRNAs, two separate risk scores were constructed for CC and CO approaches - one for OS and another for RFS. In order to dichotomize the samples into two risk groups – low-risk and high-risk groups, receiver operating characteristics curve (ROC) was used to determine the optimal cut-off point for dichotomization. Risk score was now considered as a dichotomous variable and was subjected to univariate and multivariate cox proportional hazards regression model to investigate whether the risk score would emerge as a potential independent prognostic factor. The following variables were considered as potential confounders: age at diagnosis (continuous variable), tumor stage (I, II vs III, IV), tumor grade (high vs. low) and TNBC status (Luminal vs. TNBC). The final multivariate model included those variables which were significant with $p < 0.05$. Luminal A, Luminal B and Luminal HER2 were collectively called as Luminal group. Kaplan Meier plots were used for assessing the median survival function between the two risk groups. Log-rank tests were performed to compare the survival distributions between the two risk groups. $P < 0.05$ was considered to indicate statistical significance. Hazard ratio (HR) and their corresponding 95% confidence interval (CI) are reported for the univariate and multivariate Cox' regression model.

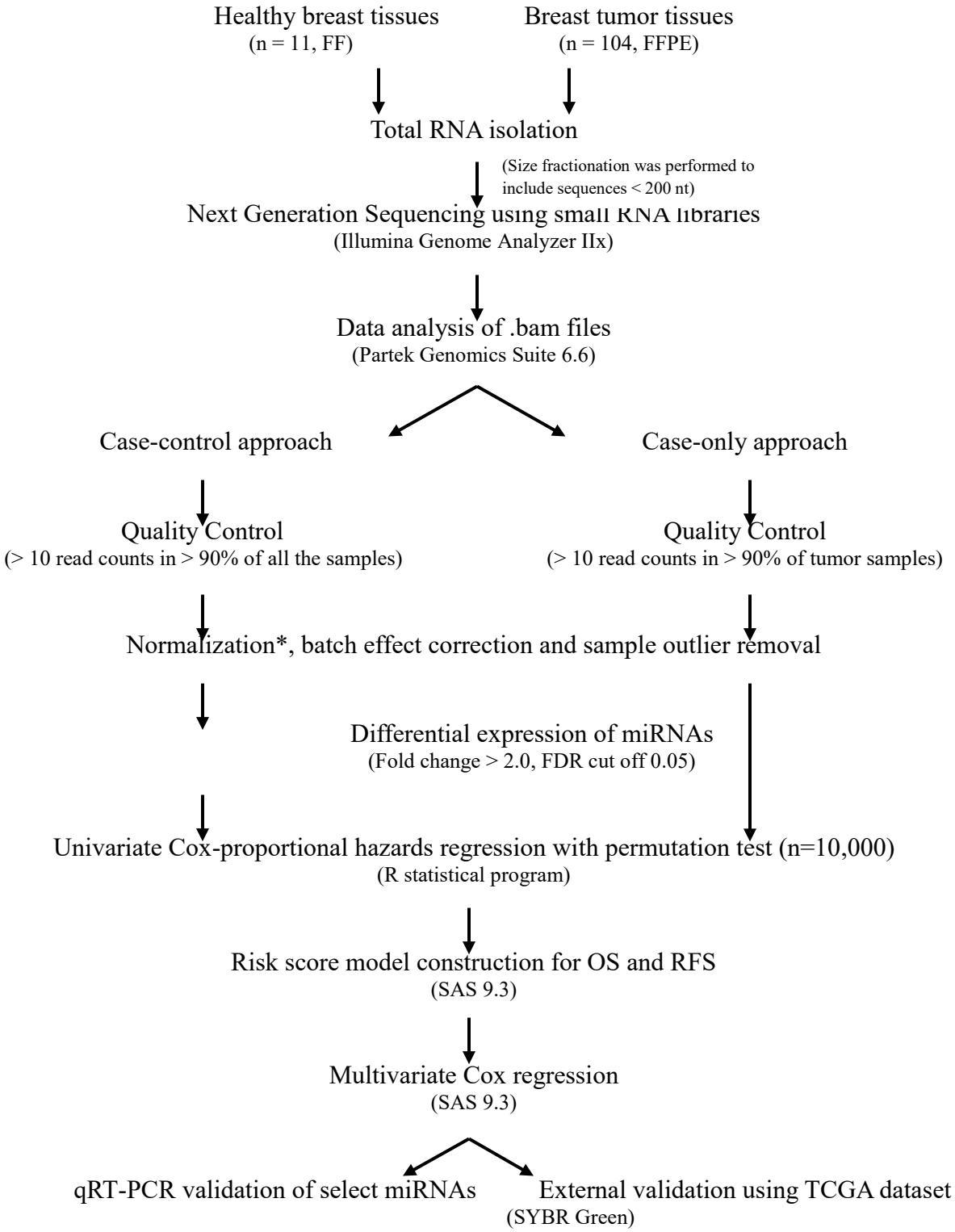
2.6.2 External validation cohort

Normalized counts that were adjusted for batch effects were used for extracting the normalized values of prognostic sncRNAs identified from the discovery cohort. As

pointed out earlier, samples selected from TCGA dataset were sequenced in Illumina HiSeq, whereas the discovery set samples were sequenced in Illumina Genome Analyzer. Due to the fact that NGS platform specific differences in read counts may potentially influence the risk scores we did not adopt the risk scores and cut-off points generated in discovery set. An independent risk score was constructed using the prognostic sncRNAs and ROC was employed to dichotomize the sample into two risk groups. Univariate, and multivariate Cox proportional hazards regression analysis was performed using the following variables: age at diagnosis (continuous variable), tumor stage (I, II vs. III and IV), TNBC status. TCGA dataset lacks information on tumor grade. However, tumor grade did not influence the multivariate analysis even in the discovery set (data not shown). Therefore I reasoned that lack of information on grade in the TCGA data set may not influence the study findings. External validation was carried out only for sncRNAs associated with OS. Since the numbers of recurrences were minimal in the external dataset, sncRNAs associated with RFS (identified from the discovery cohort) could not be validated.

An overall workflow of this study is outlined in Figure 2.1. Methods outlined in this section are common for all the sncRNAs and more specific methods are elaborated in the subsequent chapters.

Figure 2.1 Overall Workflow of the study



FF = Fresh Frozen; FFPE = Formalin Fixed Paraffin Embedded; Normalization* = Reads per kilobase per million (RPKM); FDR = False Discovery Rate; OS = Overall Survival; RFS = Recurrence Free Survival

2.7 References

1. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Guidelines: REporting recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer 2005;41:1690-1696.

2. Dobbin KK, Simon RM. Sample size planning for developing classifiers using high-dimensional DNA microarray data. Biostatistics 2007;8:101-117.

3. Dobbin KK, Zhao Y, Simon RM. How large a training set is needed to develop a classifier for microarray data? Clin Cancer Res 2008;14:108-114.

4. Bernstein L, Lacey,James V.,,Jr. Receptors, associations, and risk factor differences by breast cancer subtypes: Positive or negative? J Natl Cancer Inst 2011;103:451-453.

5. Sapkota Y, Ghosh S, Lai R, Coe BP, Cass CE, Yasui Y, et al. Germline DNA copy number aberrations identified as potential prognostic factors for breast cancer recurrence. PLoS ONE 2013;8:1-11.

6. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010;28:1684-1691.

7. Tam S, de Borja R, Tsao M, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest 2014;94:350-358.

8. Kolbert CP, Feddersen RM, Rakhshan F, Grill DE, Simon G, Middha S, et al. Multi-platform analysis of MicroRNA expression measurements in RNA from fresh frozen and FFPE tissues. PLoS ONE 2013;8:1-9.

9. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25-R25.

10. Wilson N. The miRNA registry. Nature Reviews Genetics 2004:244.

11. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006;34:D140-D144.

12. Sai Lakshmi S, Agrawal S. piRNABank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173-D177.

13. Chan PP, Lowe TM. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 2009;37:D93-D97.

14. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl genomes 2016: More genomes, more complexity. Nucleic Acids Res 2016;44:D574-D580.

15. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nature Methods 2008;5:621-628.

16. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010;11:733-739.

17. Xiong J, Bing Z, Su Y, Deng D, Peng X. An integrated mRNA and microRNA expression signature for glioblastoma multiforme prognosis. PLoS ONE 2014;9:1-8.

18. Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

3 Profiling of small non-coding RNAs from human breast tissues – normal and malignant

3.1 Small non-coding RNAs as global regulators of gene expression

A cell's phenotype is largely determined by the expression of several proteins. Proteins, as we know are translated from messenger RNAs (mRNAs) and we have observed dysregulation of proteins and their corresponding mRNAs in several diseased conditions, when compared to their normal counterparts [1]. Although mRNAs are placed a step higher in the hierarchy of protein expression regulation, there are molecules such as small non-coding RNAs (sncRNAs) that may be placed a step even higher than the mRNAs and a comprehension of these molecules may in turn enable efficient regulation of proteins. It is therefore vital to identify these molecules and understand their behavior in a diseased condition. SncRNAs are molecules which are less than 200 nt in length that serve a multitude of functions ranging from gene regulation to splicing to protein synthesis. Several classes of RNAs such as miRNAs, piRNAs, snoRNAs, tRNAs etc., are encompassed within the sncRNA family [2]. The class of sncRNAs is ever expanding with the discovery of newer molecules, along with our understanding on the roles of these molecules. miRNAs have largely been studied as regulators of gene expression. RNA world is predominantly governed by base-pair complementarity and miRNAs are a classic example of such an interaction [3-5]. The roles of other sncRNAs are being studied from different angles. For instance, piRNAs are well studied for their role in germline development and maintenance [6,7], tRNAs are crucial for protein synthesis and snoRNAs are known for their classic role in post transcriptional modifications [8]. The roles of these

other molecules are no longer restricted to their classical functions but are slowly expanding to other domains – an important domain being gene regulation. piRNAs are now known to play important roles in somatic tissues and studies are now emerging to show that piRNAs may also be classified as master regulators of gene expression which may exert its action, in a manner similar to that of miRNAs [9-12]. Several of these regulatory molecules (miRNAs and piRNAs) are also known to be embedded within slightly larger molecules such as snoRNAs [13-16] and tRNAs [17]. Recent studies have also identified the emergence of these regulatory molecules from the processing of snoRNAs and tRNAs. Thus, there is an indirect contribution of snoRNAs and tRNAs to gene regulation and a more direct role in gene regulation is an active and emerging area of research. For instance, dysregulation of snoRNAs have led to a change in the cell's phenotype [18-20], which may imply an indirect mechanism of snoRNAs in gene regulation. Also, tRNAs have been found to play a key role in activation of protein kinase GCN2 [21,22], hinting at the possibility of gene expression regulation by tRNAs. With all these insights, it may now be possible to call the sncRNA family as global regulators of gene expression. Identifying molecules that are dysregulated in a diseased condition and the genes that are regulated by these molecules may give us clues on the overall understanding of the mechanisms contributing to a condition and may also serve as possible therapeutic targets.

3.2 Objectives

Specific objectives of this chapter are to profile and identify differentially expressed (i) miRNAs, (ii) piRNAs, (iii) tRNAs and (iv) snoRNAs from human breast tissues (reduction mammoplasty vs. tumor).

3.3 Methods

Methods on isolating RNA, sequencing of sncRNAs and data analysis has been explained in detail in sections 2.3, 2.4 and 2.5.

3.4 Results

3.4.1 Descriptive analysis of sequencing data

A total of 10,016,964 reads and 164,237,348 reads were detected from normal and tumor samples, respectively. Of these, approximately 50% of the reads were retained in the normal samples after adapter trimming and 59% of the reads were retained in the tumor samples. Of the reads retained after adapter trimming, 4,255,616 reads from normal tissues and 84,240,355 reads from tumor tissues were aligned to human genome (hg19). The overall read length distribution of the aligned reads showed a size range from 17 to 27nt (since only reads with 17-27 nt length were retained), with a peak observed in 22 nt length, which corresponds to the average length of mature miRNAs (Figure 3.1). These aligned reads were mapped to four classes of small non-coding RNAs – miRNAs, piRNAs, snoRNAs and tRNAs. Table 3.1 summarizes these findings from normal and tumor breast tissues. While mature miRNAs (21 nucleotides) were adequately covered, the longer length transcripts from other sncRNAs may be potentially under-represented

due to the limitation in the read length specified in the sequencing protocol. However, the annotated RNAs may include RNAs representing different transcript lengths (actual length of the RNAs, as given in respective databases). For instance, piRNAs discovered so far range from 25-32nt. Even though the sequencing protocol captured only reads up to 27 nt, these aligned well with several of the longer piRNAs up to 32 nt. Figure 3.2 indicates the number of piRNAs identified under each transcript length.

Figure 3.1 Overall read length distribution of aligned reads

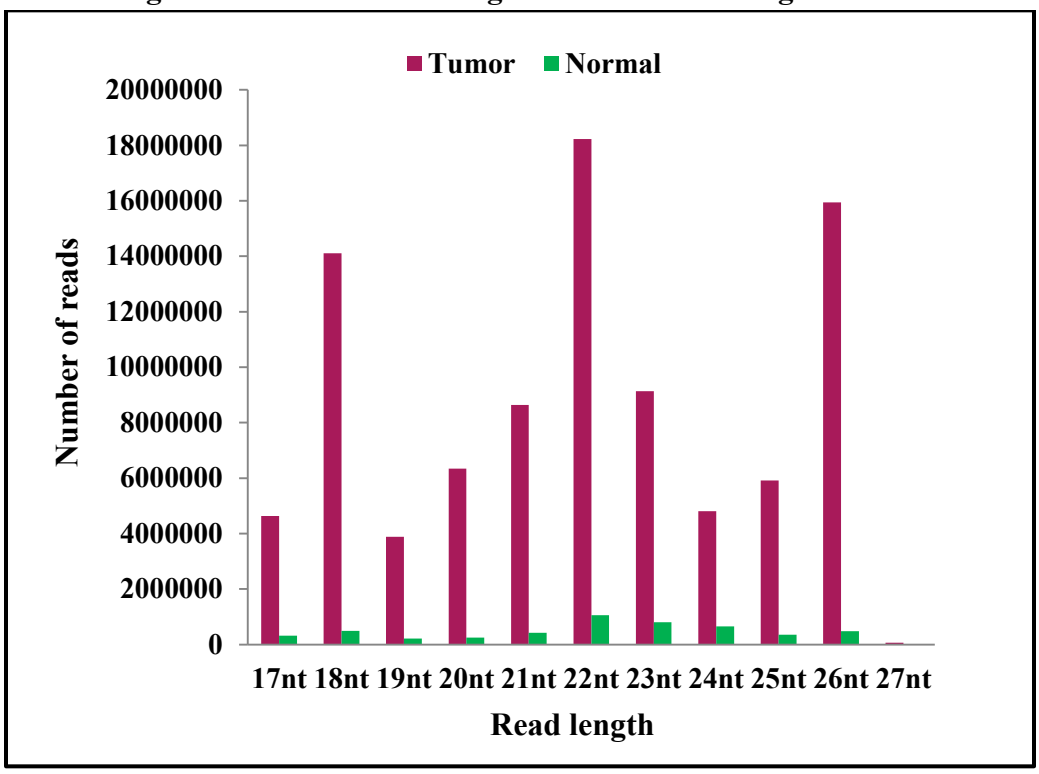
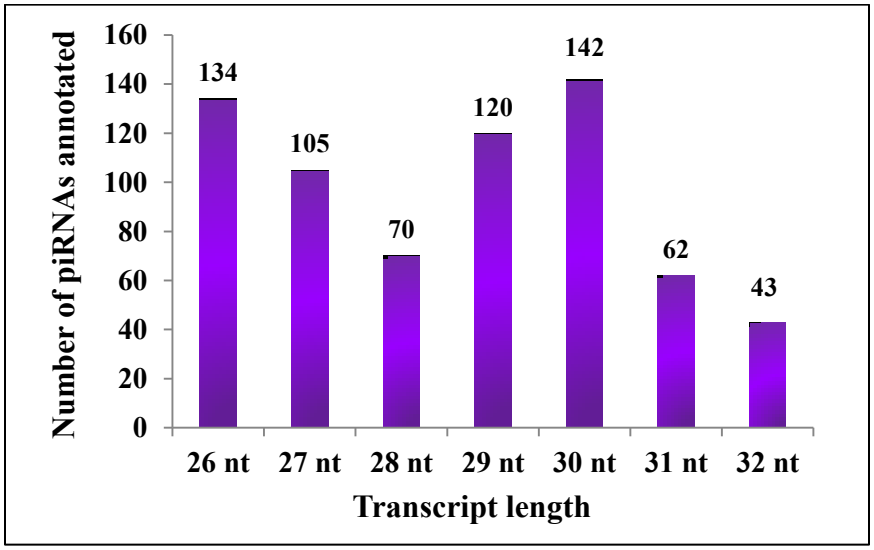


Table 3.1 Descriptive statistics of NGS data

Parameter	Normal (n=11)	Tumor (n=103)
Total reads	10,016,964	164,237,348
Reads retained after adapter trimming	5,060,588	97,204,377
Number of aligned reads	4,255,616	84,240,355
Number of unaligned reads	804,972	12,964,022
Reads mapping to miRNAs	1,174,977	24,344,516
Reads mapping to piRNAs	307,485	3,899,537
Reads mapping to tRNAs	124,352	8,122,670
Reads mapping to snoRNAs	163,459	1,447,469

Figure 3.2 Transcript length of the piRNAs identified from normal and tumor tissues



3.4.2 Exploratory analysis of sequencing data

Reads mapping to different classes of sncRNAs were annotated to different sncRNA IDs. The dataset of each of the classes of sncRNAs was analyzed for potential batch effects and was also interrogated for potential sample outliers. The datasets were first filtered for read counts – only those RNAs with a minimum of 10 read counts (raw counts) in at least 90% of the samples (normal and tumor samples inclusive) were

retained. The datasets were normalized (including all the RNAs) and were then filtered for read counts (based on the raw counts). The normalized counts of filtered RNAs were quantified and corrected for potential batch effects using ANOVA model. Along with this technical variation, biological variation was also captured for comparison and to ensure that only the technical variations are removed. Mean F ratios of both the biological and technical variations (in this case, batch) were measured and were compared to the mean F ratio of error. Those variations having a mean F ratio above the error bar (which is always 1) are considered as a source of variation to the dataset. Since the variation from biological source (tumor and normal tissues) is expected, this variation was not adjusted for. All the four sncRNAs had measurable amounts of batch effects (Table 3.2, Appendix figures 9.1.1 – 9.1.4). The datasets were adjusted for batch effects and the mean F ratio values dropped to zero, value less than the error ratio of 1, indicating the dataset has been corrected for batch effects (Appendix figures 9.1.1 – 9.1.4). The datasets, after adjusting for batch effects, were interrogated for potential sample outliers and for each of the datasets; unique/overlapping samples from different sncRNA datasets were identified as outliers (Table 3.2). These samples were removed and the datasets (without outliers) were reloaded, normalized, filtered for read counts, adjusted for batch effects and analyzed for differential expression of sncRNAs between normal and tumor tissues.

Table 3.2 Results of exploratory analysis of sequencing data

Dataset analyzed	Mean F ratio of batch (before batch correction)	Number of sample outliers
miRNA	3.81	1
piRNA	10.02	1
snoRNA	19.73	3
tRNA	14.48	1

3.4.3 Identification of differentially expressed sncRNAs

The annotated reads of each class of sncRNAs were filtered for read counts, as explained above. The dataset normalized and filtered for read counts and adjusted for batch effects were interrogated for differential expression of sncRNAs. All sncRNAs exhibiting fold change ≥ 2.0 and FDR ≤ 0.05 were classified as differentially expressed sncRNAs. The number of DE RNAs identified from each dataset and the steps involved in arriving at the DE list are summarized in Table 3.3. The list of DE sncRNAs is given in the Appendix table 9.1. Further, unsupervised hierarchical clustering of samples for each class of sncRNAs revealed clear separation of tumor and normal samples; illustrating that the samples are differentiated by the relative expression of a common set of sncRNAs rather than by unique sncRNAs (Appendix figures 9.2.1 – 9.2.4).

Table 3.3 Identification of differentially expressed sncRNAs

Analysis step	miRNAs	piRNAs	tRNAs	snoRNAs
Number of RNAs annotated	1423	676	572	768
RNAs retained after filtering	126	42	148	88
Differentially expressed RNAs	80	25	76	40
Up-regulated RNAs	48	17	76	9
Down-regulated RNAs	32	8	0	31

3.5 Discussion

The class of non-coding RNAs are broadly categorized into small and long non-coding RNAs, based on their size [2]. The focus of this study is on small non-coding RNAs which are < 200 nt in length. Several classes of RNAs fall under the category of sncRNAs and these include miRNAs, piRNAs, snoRNAs, tRNAs, etc [2]. The discovery of miRNAs, their roles in various key cellular mechanisms and [23-26] their potential as

promising biomarkers for cancer [27-35] and other disease phenotypes has [36] revolutionized our understanding of non-coding RNAs [37], which were for a long time largely ignored in the mistaken belief that they served no meaningful roles in somatic cells. As we understand today, non-coding RNAs serve a wide repertoire of functions, with substantial roles in gene regulation and knowledge in this area is growing exponentially. Given the widespread biological functions of small RNAs, comprehensive profiling and comparison of the expression patterns of these RNAs in a case-control study would serve as a cue for understanding the consequences of their abnormal expression patterns in the diseased state.

3.5.1 Technical considerations

An inert problem in high throughput techniques is the occurrence of batch effects, the inclusion of which would result in spurious associations [38]. Although reports have suggested that the problem of batch effects is less encountered in NGS data, yet the NGS data is not free from it [39]. I acknowledge this fact as batch effects were observed and quantified in this study. However, I used ANOVA model to adjust for batch effects and the same was confirmed when the mean F ratios of batch effects were found to zero, indicating the absence of technical variation. Moreover, PGS calculates the p-value for batch effects too, when we analyze for differential expression of RNAs. A p-value of 1 was found after batch effects correction (which was $p < 1$ before correction) for each studied small RNA class, indicating negation of batch effects from the datasets.

3.5.2 microRNAs

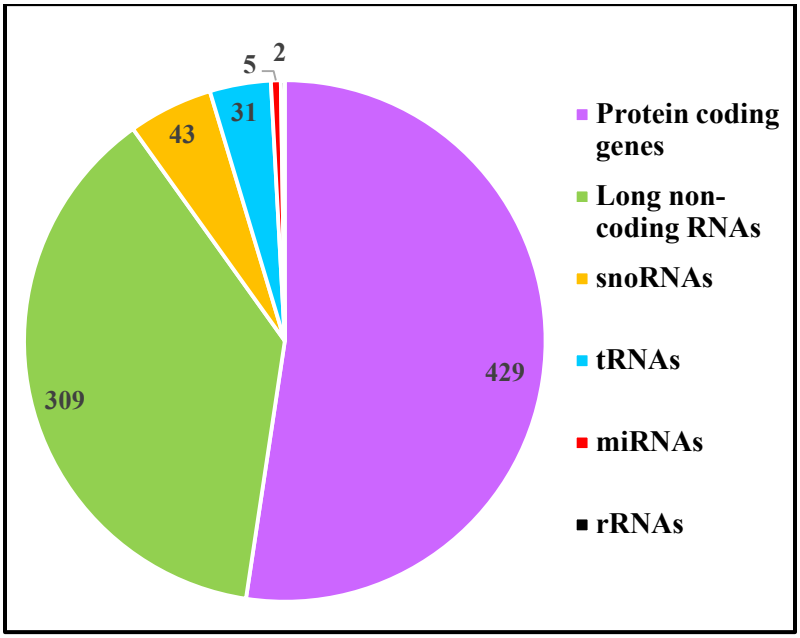
In recent years, microRNAs have gained prominence as valuable biomarkers for several cancer types, including BC. Although considerable progress has been made in this field, clinical application of these RNAs as prognostic markers has not yet been possible because of the generation of different signatures by different studies with only a small number of overlapping molecules. This discrepancy may be attributed to several reasons, the primary being the use of different profiling platforms. While there are ~2,588 miRNAs identified so far, as reported in the miRBase, only a few hundreds have been captured on microarray or qRT-PCR platforms, of which even fewer have been detected in breast tissues, since miRNAs are tissue specific. On the other hand, NGS profiling of the entire miRNAome, including even the less abundant ones, can now be used to probe the larger repertoire, which was evident from this study. Approximately, breast tissue specific miRNAs comprise 55% (n = 1,423) of the total miRNAs (n = 2,588 annotated thus far) and these were captured from the 11 normal breast tissues and 103 breast tumor tissues used for the study. 80 miRNAs were also found to be DE, with 48 up-regulated and 32 down-regulated miRNAs. Of these 80 miRNAs, I interrogated the direction of expression of a subset of miRNAs showing prognostic significance (see chapter 3) with the published miRNA profiling studies in breast cancer and other cancer types. All of the tested miRNAs exhibited excellent concordance [40-53]. These findings gave the confidence to further mine the NGS data to interrogate other sncRNAs.

3.5.3 Piwi-interacting RNAs

A new class of non-coding RNAs called piRNAs was discovered in mouse testes in 2006 [54-57]. They were found to be involved in maintaining genome stability by

regulating the expression of transposons in germ cells [6,7], and for a long time, their roles beyond germ cells remained uncertain; however, with increasing focus on these molecules, their occurrence in somatic cells has been observed and their functional roles in somatic cells are beginning to be uncovered [9-12]. Using a sequencing platform to profile piRNAs, we observed the presence of 676 piRNAs in breast tissues, confirming their existence in somatic tissues. In contrast to their occurrence as clusters in germ cells, they have been observed mapping to known transcripts in somatic cells [58]. In breast tissues alone, we noted that around 85% (576 of the 676 total piRNAs profiled in our dataset) of the piRNAs mapped to exons and introns of known protein coding and non-coding transcripts. Since piRNAs abundantly map to known genes, it remains to be determined if they are dependent on the host gene's promoter for their transcription or if they carry their own promoter. A few of the piRNAs also mapped to other non-coding RNA classes such as miRNAs, tRNAs and snoRNAs (Figure 3.3). These piRNAs shared genomic co-ordinates with the other classes of ncRNAs, either fully or partially. In case of snoRNAs and tRNAs, several of the piRNAs were found to be completely embedded within the larger tRNAs (n = 31) and snoRNAs (n = 38).

Figure 3.3 Expression of piRNAs in breast tissues



Filtering of these 676 piRNAs for read counts and differential expression analysis identified 25 piRNAs with FC > 2 and FDR ≤ 0.05. The contribution of piRNAs to tumorigenesis is beginning to be understood and characterization of the identified piRNAs is required to gain insights into their specific functions.

3.5.4 Transfer RNAs

tRNAs are among the most abundant molecules present in cells, especially in a metabolically active disease setting such as cancer, indicating higher rate of protein synthesis in these cells [59]. Despite their abundance, they have received less attention as biomarkers mainly due to the complexities involved in developing a profiling platform [60]. The extensive modifications that a tRNA undergoes during maturation and the complicated structure of mature tRNA have deterred the development of a profiling platform as these structural intricacies interfere with reverse transcription and hybridization protocols [60,61]. In 2006, however, Dittmar et al. developed a microarray

method to profile tRNAs on a genome-wide scale. These microarray protocols could also distinguish between tRNA isoacceptors [60]. Although this is a major leap in the field of tRNA profiling, the method requires custom-made arrays (which may not tend to be cost-effective for large scale profiling experiments) and has limited dynamic range for quantification [62]. However, a recent report by Meng et al. confirmed that a wider range of molecules such as the class of sncRNAs, including tRNAs can be profiled using NGS and from clinically archived specimens, preserved as FF or FFPE tissue blocks [63]. tRNAs, however posed challenges even in high-throughput sequencing platforms, mainly due to their compact tertiary structure and the presence of post-transcriptional modifications. This limits the adapter binding efficiency and reverse transcription, both of which are needed to generate libraries and to perform sequencing experiments, resulting in the generation of truncated sequences from a large subset of tRNAs [60,61]. Despite this difficulty in tRNA sequencing, I observed a higher number of reads aligning to tRNAs (n = 8,247,022) when compared to piRNAs (n = 4,207,022), snoRNAs (n = 1,610,928) and snRNAs (n = 435,276) but only secondary to miRNAs (n = 25,003,223). This observation may be attributed to the abundance of tRNAs in the cells and tissues and despite the challenges in the sequencing of tRNAs with high secondary structure or base modifications.

Illumina sequencing protocols are still emerging to overcome the inherent limitations described above. I expect a higher number of reads than reported, if the sequencing limitations are overcome. Also, the small RNA sequencing protocol using TruSeq Small RNA library preparation kit has been designed to capture RNAs possessing 5’ phosphate and 3’ hydroxyl group. Ligation of adapters occurs at the ends of RNAs that

possess these modification but these adapters can also ligate to other RNAs, albeit inefficiently, which may therefore contribute to the lower abundance of reads mapping to tRNAs. Also, given the read length restriction adopted in our sequencing analysis (17 – 27 nt), it may not have been feasible to capture full length tRNAs. Therefore the reads captured in this study may likely be the fragments of tRNAs but it is not certain if these fragments are representative of mature full length tRNAs or if they represent actual physiological products (identified as tRFs). Nevertheless, results from this study and other studies [64] confirm that tRNA sequences can be accurately captured using small RNA sequencing and the reads captured from our study represent the known tRNAs identified and annotated to-date across all chromosomes (Table 3.5). Till date, in the human genome, 625 tRNA genes (including pseudogenes) have been identified [65], of which 571 were profiled in this study. tRNAs predominantly arise from chromosome 6 (n = 175), followed by chromosome 1 (n = 137). In this dataset of 571 tRNAs, I have also observed a similar pattern of distribution (Table 3.4), with 170 tRNAs arising from chromosome 6 and 132 from chromosome 1, indicating an unbiased genome wide capture of tRNAs using the NGS platform.

Table 3.4 Distribution of tRNAs profiled in the breast tissues

Chromosome	Number of tRNAs identified in human genome	Number of tRNAs identified in our dataset
1	139	132
2	28	22
3	12	10
4	5	4
5	24	21
6	175	170
7	26	25
8	14	9
9	8	6
10	6	3
11	19	16
12	16	14
13	7	7
14	23	22
15	11	10
16	34	33
17	42	38
18	4	4
19	14	12
20	7	4
21	2	1
22	1	1
X	7	6
Y	1	1

76 tRNAs were up-regulated in breast tumor samples, which independently confirms the findings from Pavon-Eternod et al. who also reported an overall up-regulation of tRNAs in breast cancer [66]. In the current study, tRNAs coding for 14 amino acids (Arg, Asn, Asp, Gln, Glu, Gly, His, Leu, Lys, Met, Phe, Ser, Val and SeC(e)) clearly showed high DE (FC > 2 and FDR $\leq$ 0.05). The global up-regulation of tRNAs may be attributed to the high metabolic activity of the cancer cell requiring higher rates

of protein synthesis and tRNAs per se may serve diverse non-canonical roles in the cell. Although this phenomenon (global up-regulation of tRNAs) is observed for BC, it remains to be seen if similar patterns of expression exists in other cancer tissues.

In the study by Pavon-Eternod, the authors also pointed out the differences in tRNA isoacceptor levels, correlating with the codon preferences of genes involved in tumorigenesis [66]. Although we did not focus on codon-isoacceptor correlations, we did observe differences in tRNA isoacceptor levels for specific amino acids (Appendix table 9.2) in our study, which may correlate with the codon preferences of the genes. For instance, $tRNA^{Arg(TCT)}$ and $tRNA^{Arg(CCG)}$ were expressed in higher amounts when compared to $tRNA^{Arg(CCT)}$. Similarly, $tRNA^{Leu(CAG)}$, $tRNA^{Leu(CAA)}$, and $tRNA^{Leu(TAA)}$ were over expressed, when compared to $tRNA^{Leu(AAG)}$ and $tRNA^{Leu(TAG)}$. tRNAs coding for Gln, Glu and Val also showed preferential expression of certain isoacceptors. In contrast, expression changes of isoacceptors for tRNAs coding for Ser, Gly and Lys remained invariant. In this study, I have observed preferential expression of certain isoacceptors over the others. However future studies are necessary to identify the codon preferences of genes involved in breast tumorigenesis and subsequently correlate it with the tRNA isoacceptor levels. Studies of this kind may further help us understand how tRNAs may directly be involved in breast tumor development.

Overall, this is the largest study to attempt comprehensive profiling of tRNAs. Although sequencing platform imposed technical difficulties in profiling tRNAs, this study still stands as a proof of principle experiment to demonstrate that tRNAs can be captured, even at the iso-acceptor levels through traditional sequencing protocols

3.5.5 Small nucleolar RNAs (snoRNAs)

One of the lesser studied class of sncRNAs is the small nucleolar RNAs which plays a major role in post transcriptional modifications of other RNAs and in ribosomal RNA biogenesis [8,67]. This RNA also acts a hub to promote interplay of other RNA molecules. Although this class of RNAs is not known to be directly involved in gene regulation, it may influence gene regulation in an indirect way. One of the emerging fields of research focuses on the processing of snoRNAs to other regulatory sncRNAs such as miRNAs [13-15] and piRNAs [16], thus addressing the role of snoRNAs in gene regulation. The clinical relevance of snoRNAs is also slowly coming to the fore as potential biomarkers for various cancer types [20,68,69]. Therefore there is a pressing need to understand these molecules from different perspectives. Although snoRNAs may also be captured using a microarray platform, NGS serves as a better choice (for reasons explained in 1.3.3) to enable comprehensive profiling of snoRNAs. So far genome wide profiling of snoRNAs has not been a common sight in literature and this dataset, which has captured 768 snoRNAs is the largest dataset interrogated for any cancer type. snoRNAs are well known to be embedded within the intronic regions of protein coding genes [70]. I also observed a similar trend in the genomic location of snoRNAs, where a majority of the snoRNAs profiled (449 out of 768), mapped to the intronic regions of the protein coding genes. A total of 40 snoRNAs were also found to be dysregulated, suggesting their possible involvement in breast tumorigenesis. However, their exact roles in breast tumorigenesis remain to be delineated.

Overall, I was able to profile a total of 768 snoRNAs from breast tissues. Given the sequencing protocol adopted in this study (36 cycles single end protocol) with read

lengths ranging between 17 and 27 nucleotides, it is highly likely that the 768 snoRNAs may not represent the entire snoRNAome. Since full-length snoRNAs have a minimum length of 60 nucleotides, the identified snoRNAs may actually be fragments of snoRNAs. Increasing the number of sequencing cycles may help identify additional snoRNAs and confirm the origins of the profiled fragments.

3.6 Conclusions

In literature, especially in breast cancer literature, it is common to see studies focussing exclusively on miRNAs. In this study, I have attempted a comprehensive profiling of four sncRNAs, exhibiting some percentage of commonality in their contribution to gene regulation. While we now know substantially about miRNAs, our knowledge on other sncRNAs is still rudimentary and such a comprehensive study may cater to better understanding about these molecules and of the phenotype under study. When compared to other published studies, the number of RNAs annotated and interrogated under each class of sncRNAs is the largest in this study. It is also one of the rare studies to have complete clinical annotation and long follow-up period for all the samples, which is one of the critical determinants of a good biomarker study. While this chapter has focussed exclusively on genome-wide profiling of sncRNAs, subsequent chapters will delve into the clinical significance of the four classes of sncRNAs and identify potential biomarkers for breast cancer.

3.7 References

1. CIendening JW, Pandyra A, Boutros PC, Ghamrasni SE, Khosravi F, Trentin GA, et al. Dysregulation of the mevalonate pathway promotes transformation. Proc Natl Acad Sci U S A 2010;107:15051-15056.

2. Ronnau C, Verhaegh GW, Luna-Velez M, Schalken JA. Noncoding RNAs as novel biomarkers in prostate cancer. Biomed research international 2014;2014:591703.

3. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 2004;116:843-854.

4. Bartel DP. Review: MicroRNAs. genomics, biogenesis, mechanism, and function. Cell 2004;116:281-297.

5. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 2004;116:855-862.

6. Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the drosophila melanogaster flamenco gene. Genetics 2004;166:1313-1321.

7. Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev 1998;12:3715-3727.

8. Weinstein LB, Steitz JA. Guided tours: From precursor snoRNA to functional snoRNP. Curr Opin Cell Biol 1999;11:378-384.

9. Peng L, Song L, Liu C, Lv X, Li X, Jie J, et al. piR-55490 inhibits the growth of lung carcinoma by suppressing mTOR signaling. Tumour Biol 2015;37:2749-2756.

10. Watanabe T, Lin H. Posttranscriptional regulation of gene expression by piwi proteins and piRNAs. Mol Cell 2014;56:18-27.

11. Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, et al. Identification of novel piRNAs in bladder cancer. Cancer Lett 2015;356:561-567.

12. Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-9910.

13. Martens-Uzunova E, Olvedy M, Jenster G. Beyond microRNA - novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett 2013;340:201-211.

14. Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, et al. Article: A human snoRNA with MicroRNA-like functions. Mol Cell 2008;32:519-528.

15. Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res 2011;39:675-686.

16. Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, et al. A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. Nucleic Acids Res 2015;43:10474-10491.

17. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 2013;110:1404-1409.

18. Mei Y, Liao J, Shen J, Yu L, Liu B, Liu L, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. Oncogene 2012;31:2794-2804.

19. Chu L, Su MY, Maggi L,Jr, Lu L, Mullins C, Crosby S, et al. Multiple myeloma-associated chromosomal translocation activates orphan snoRNA ACA11 to suppress oxidative stress. J Clin Invest 2012;122:2793-2806.

20. Okugawa Y, Toiyama Y, Toden S, Mitoma H, Nagasaka T, Tanaka K, et al. Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. Gut 2015.

21. Wek SA, Zhu S, Wek RC. The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. Mol Cell Biol 1995;15:4497-4506.

22. Dong J, Qiu H, Garcia-Barrio M, Anderson J, Hinnebusch AG. Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. Mol Cell 2000;6:269-279.

23. Sayed D, Abdellatif M. MicroRNAs in development and disease. Physiol Rev 2011;91:827-887.

24. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res 2005;65:6029-6033.

25. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A 2005;102:13944-13949.

26. Chen C, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science 2004;303:83-86.

27. Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics 2015;16:735-735.

28. Rabinowits G, Gerçel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal MicroRNA: A diagnostic marker for lung cancer. Clin lung cancer 2009;10:42-46.

29. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008;18:997-1006.

30. Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

31. Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-microRNA signature. Gut 2010;59:579-585.

32. Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-1184.

33. Cuk K, Zucknick M, Heil Jö, Madhavan D, Schott S, Turchinovich A, et al. Circulating microRNAs in plasma as early detection markers for breast cancer. Int J Cancer 2013;132:1602-1612.

34. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 2013;19:4477-4487.

35. Kleivi Sahlberg K, Bottai G, Naume B, Burwinkel B, Calin GA, Børresen-Dale A, et al. A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. Clin Cancer Res 2015;21:1207-1214.

36. Miller BH, Zeier Z, Xi L, Lanz TA, Deng S, Strathmann J, et al. MicroRNA-132 dysregulation in schizophrenia has implications for both neurodevelopment and adult brain function. Proc Natl Acad Sci U S A 2012;109:3125-3130.

37. Berindan-Neagoe I, Monroig PD, Pasculli B, Calin GA. MicroRNAome genome: A treasure for cancer diagnosis and therapy. CA-A Cancer journal for clinicians 2014;64:311-336.

38. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010;11:733-739.

39. Tam S, de Borja R, Tsao M, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest 2014;94:350-358.

40. Wang L, Zhu M, Ren A, Wu H, Han W, Tan R, et al. A ten-MicroRNA signature identified from a genome-wide MicroRNA expression profiling in human epithelial ovarian cancer. PLoS ONE 2014;9:1-8.

41. Scarpati GDV, Calura E, Di Marino M, Romualdi C, Beltrame L, Malapelle U, et al. Analysis of differential miRNA expression in primary tumor and stroma of colorectal cancer patients. BioMed Research International 2014:1-8.

42. Wu X, Pu X, Zhao Z, Zhao Y, Kang X, Wu W, et al. The expression profiles of microRNAs in kaposi's sarcoma. Tumor Biol 2015:437.

43. Kodahl AR, Lyng MB, Binder H, Cold S, Gravgaard K, Knoop AS, et al. Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: A case control study. Molecular Oncology 2014;8:874-883.

44. Wang W, Liu L, Li G, Chen Y, Li C, Jin D, et al. Combined serum CA19-9 and miR-27a-3p in peripheral blood mononuclear cells to diagnose pancreatic cancer. Cancer Prev Res (Phila) 2013;6:331-338.

45. Rani S, Gately K, Crown J, O'Byrne K, O'Driscoll L. Global analysis of serum microRNAs as potential biomarkers for lung adenocarcinoma. Cancer Biol Ther 2013;14:1104-1112.

46. Kim B, Hong SW, Kim A, Choi SH, Yoon SO. Prognostic implications for high expression of oncogenic microRNAs in advanced gastric carcinoma. J Surg Oncol 2013;107:505-510.

47. Teixeira AL, Ferreira M, Silva J, Gomes M, Dias F, Santos JI, et al. Higher circulating expression levels of miR-221 associated with poor overall survival in renal cell carcinoma patients. Tumor Biol 2014;35:4057-4066.

48. Li J, Wang Y, Yu W, Chen J, Luo J. Expression of serum miR-221 in human hepatocellular carcinoma and its prognostic significance. Biochem Biophys Res Commun 2011;406:70-73.

49. Falkenberg N, Anastasov N, Rappl K, Braselmann H, Auer G, Walch A, et al. MiR-221/-222 differentiate prognostic groups in advanced breast cancers and influence cell invasion. Br J Cancer 2013;109:2714-2723.

50. Cai J, Guan H, Fang L, Yang Y, Zhu X, Yuan J, et al. MicroRNA-374a activates wnt/ß-catenin signaling to promote breast cancer metastasis. J Clin Invest 2013;123:566-579.

51. Rothé F, Ignatiadis M, Chaboteaux C, Haibe-Kains B, Kheddoumi N, Majjaj S, et al. Global MicroRNA expression profiling identifies MiR-210 associated with tumor proliferation, invasion and poor clinical outcome in breast cancer. PLoS ONE 2011;6:1-13.

52. Yang G, Han D, Chen X, Zhang D, Wang L, Shi C, et al. MiR-196a exerts its oncogenic effect in glioblastoma multiforme by inhibition of IκBα both in vitro and in vivo. Neuro Oncol 2014;16:652-661.

53. Li X-, Yan P-, Shao Z-. Downregulation of miR-193b contributes to enhance urokinase-type plasminogen activator (uPA) expression and tumor progression and invasion in human breast cancer. Oncogene 2009;28:3937-3948.

54. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian piwi proteins. Nature 2006;442:199-202.

55. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 2006;442:203-207.

56. Grivna ST, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20:1709-1714.

57. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science 2006;313:363-367.

58. Martinez VD, Vucic EA, Thu KL, Hubaux R, Enfield KSS, Pikor LA, et al. Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. Sci Rep 2015;5:10423-10423.

59. Dolfi SC, Chan LL, Qiu J, Tedeschi PM, Bertino JR, Hirshfield KM, et al. The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. Cancer & Metabolism 2013;1:20.

60. Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. PLoS Genetics 2006;3:2107-2115.

61. Sprinzl M, Vassilenko KS. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 2005;33:D139-D140.

62. Yan Ling JP, Abo R, Levine SS, Dedon PC. Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. Nucleic Acids Res 2014;42:e170.

63. Meng W, McElroy JP, Volinia S, Palatini J, Warner S, Ayers LW, et al. Comparison of MicroRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. PLoS ONE 2013;8:e64393.

64. Guo Y, Bosompem A, Mohan S, Erdogan B, Ye F, Vickers KC, et al. Transfer RNA detection by small RNA deep sequencing and disease association with myelodysplastic syndromes. BMC Genomics 2015;16:727.

65. Chan PP, Lowe TM. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 2009;37:D93-D97.

66. Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res 2009;37:7268-7280.

67. Filipowicz W, Pogacic V. Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol 2002;14:319-327.

68. Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkersfor non-small-cell lung cancer. Molecular Cancer 2010;9:198-207.

69. Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, et al. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. Int J Cancer 2015;136:E623-E629.

70. Dieci G, Preti M, Montanini B. Minireview: Eukaryotic snoRNAs: A paradigm for gene expression flexibility. Genomics 2009;94:83-88.

4 Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer

4.1 Introduction

The global burden of breast cancer (BC) is 1.7 million and is one of the leading causes of cancer related death among women and the most frequently diagnosed cancer in 140 of 182 countries, as per the 2012 statistics [1]. Although advancements in diagnosis, screening and awareness help identify BC at an early stage, optimal management has remained a challenge due to its histological and molecular heterogeneity [2], and varying response to therapies even within clinical subtypes of BC [3]. Identification and validation of prognostic markers that can stratify patients based on their risk for recurrence and/or death may help in optimizing therapies to improve disease outcomes and quality of life. Estrogen Receptor (ER) and Human Epidermal Growth Factor Receptor 2 (HER2) are widely being used as both prognostic and predictive markers but remain as imperfect estimators of the risk for recurrence [4]. While, messenger RNA (mRNA) signatures from global gene expression profiling have also been put forth as potential prognostic markers for BC [5-8], their utility is limited to specific clinical settings [9]. This further emphasizes the need to identify robust prognostic markers with higher sensitivity, accuracy and reproducibility.

A version of this chapter has been published. Krishnan et al., 2015. BMC Genomics. I would like to thank Jennifer Dufour and Mieke Heyns for providing technical assistance.

MicroRNAs (miRNAs, 18-25 nt) are evolutionarily conserved small non-coding RNAs that have shown promise as both diagnostic and prognostic biomarkers for several cancer types [10]. Predominantly, miRNAs behave as post-transcriptional regulators of gene expression, promoting either mRNA degradation or translation inhibition, depending upon the complementarity shared between the seed sequence of miRNAs and the corresponding 3' untranslated region of the target sequence [11-13]. However, studies have shown that they also activate gene expression [14]. Being pleiotropic (one miRNA regulating several mRNAs) and highly redundant (several miRNAs targeting one mRNA) in nature [15], the impact of miRNA dysregulation in cancer is complex and yet promising in the overall landscape of tumorigenesis and prognostication.

Although several studies have highlighted the significance of miRNAs as diagnostic [16,17] and prognostic markers for various cancers [18,19], including BC [20-23], a consensus signature has not yet been identified due to differences in the profiling platforms employed, analytical approaches implemented, sample types (e.g. adjacent normal tissues or reduction mammoplasty specimens) used for analysis and tumor heterogeneity. The majority of the studies have utilized profiling platforms such as microarray or qRT-PCR, which are limited to the detection of known targets at the time of assay development. Hybridization platforms are also burdened with the problems of cross hybridization, background signal, low sensitivity and limitations on the dynamic range of detection. These problems are now overcome by Next Generation Sequencing (NGS) platforms [24]. NGS also offers the advantage of capturing not just miRNAs but a whole repertoire of small RNAs, even those present in low abundance [25], thus enabling a comprehensive

analysis of small RNAome. However, despite several advantages offered by NGS, only few studies have utilized NGS platform to identify prognostic markers for BC [26,27].

Statistical methods implemented in a study also play a vital role in determining the reproducibility of findings in a prognostic signature. Two methods to identify prognostic markers are widely used in the published literature– the case-control (CC) approach [22,23] and the case-only (CO) approach [18,19,28]. While the former method utilizes a set of differentially expressed miRNAs for downstream analysis, the latter offers the advantage of being unbiased in selecting miRNAs for further analysis. Although each of the methods has been used in published miRNA studies, no study has analyzed a dataset using both the methods to compare and identify the best approach.

In this study, I hypothesized that relative variations in miRNA expression in tumors and/or apparently normal (non-malignant) tissues contribute to inter-individual differences in disease trajectory and eventual treatment outcomes. I profiled miRNAs from 104 breast cancers, predominantly of Luminal A and triple negative subtypes and 11 normal tissues (reduction mammoplasty specimens) using the NGS platform. The specific objectives were as follows: (i) to identify differentially expressed miRNAs in breast tissues (normal vs. tumor tissues) and (ii) to identify miRNAs as prognostic markers (outcome: Overall Survival, OS and Recurrence Free Survival, RFS) for BC and validate the signatures using an external dataset. I have identified a total of twelve miRNAs associated with OS and/or RFS for BC. Of these twelve, the prognostic significance of ten miRNAs already reported in literature for BC has been replicated. To the best of my knowledge, this is the first study to report two novel miRNAs (miR-574-3p and miR-660-5p) for BC prognosis.

4.2 Materials and methods

Details on samples used for the study as discovery and external validation cohorts have been explained in detail in section 2.2 (2.2.1 for discovery cohort and 2.2.2 for external validation cohort). Also, the methods that were followed for total RNA isolation, small RNA sequencing and sequence data analysis and survival analysis have been elaborated in 2.3, 2.4, 2.5 and 2.6, respectively.

Specific methods for analyzing miRNA data are explained below.

4.2.1 qRT-PCR validation of select miRNAs

qRT-PCR experiments were performed in collaboration with Dr. Kovalchuk's laboratory in University of Lethbridge and I analyzed and interpreted the data.

Using samples for which RNA was available following NGS, the expressions of three down-regulated miRNAs whose FC ranged from -1.3 to -5.8 and one up-regulated miRNA with FC = 12.8 was validated using qRT-PCR. This was done to exemplify the dynamic range of detection and concordance between NGS and qRT-PCR. This analysis included two representative miRNAs (miR-574-3p and miR-660-5p) that were identified to be of prognostic value and considered as novel in BC. miR-99b-5p (FC = -2.3), miR-574-3p (FC = -5.8), miR-769-5p (FC = -1.3) and miR-660-5p (FC = 12.8) were validated using miScript II RT kit (QIAGEN), miScript SYBR Green PCR kit (QIAGEN) and their corresponding miScript Primer Assays according to manufacturers' instructions. All assays were performed in triplicates and human RNU6-2 (QIAGEN) served as the loading control. Fold-expression changes of miRNAs were calculated using the $2^{-\Delta\Delta Ct}$ method [29].

4.2.2 Breast tumor transcriptome analysis (mRNA) and identification of targets for miRNAs

Messenger RNA (mRNA) expression datasets generated previously (GEO accession ID GSE22820) using Agilent microarray platform were available in-house (collaboration with Dr. John Mackey) for 176 tumor samples and 10 normal (reduction mammoplasty) samples. Of these, the raw files were available for 149 tumor samples and for all the normal samples. Since HER2+ samples were not utilized in the NGS experiment for miRNAs, I removed HER2+ samples from the gene expression data as well, leaving 141 tumor samples and 10 normal samples for the analysis. Seventeen tumor samples matched with the tumor samples in the discovery cohort used for NGS experiment. Raw intensity values were Quantile normalized and log 2 transformed, and one-way ANOVA was performed to identify DE genes with FC > 2.0 and FDR cut-off of 0.05 (PGS 6.6). I performed two experiments to choose the right sample set for identifying gene targets. First experiment was carried out using all the tumor samples (n = 141) and normal samples (which will henceforth be called as "all sample dataset") and the second experiment was performed using the 17 matched tumor samples and all the normal samples (which will henceforth be called as "matched sample dataset"). DE mRNAs were identified from both the experiments.

I first predicted mRNA targets for miRNAs associated with OS and RFS *in silico* using TargetScan database (Version 6.2) (http://www.targetscan.org/). The targets thus obtained were overlapped with DE mRNAs generated from the in-house dataset (from both the experiments). The benefit of using mRNA datasets from breast tissues is that they act as a proxy for functional validation of mRNA targets identified by the *in silico*

prediction algorithm. I did not restrict identification of targets only to those exhibiting inverse relationships with miRNAs (such as up-regulated miRNAs were matched with down-regulated genes and vice versa), but any correlation of miRNA to mRNA was captured since the miRNA-mRNA interactions are more complex than the direct regulation of targets by miRNAs. Gene ontology (GO) terms were identified for targets of every miRNA separately using DAVID bioinformatics tools v6.7 (http://david.abcc.ncifcrf.gov/) [30]. Only clusters with enrichment scores (ES) $\geq$ 1.3 [30] were used to identify specific GO terms related to cancer with $p<0.05$.

4.3 Results

Initial results on profiling of miRNAs and identification of DE miRNAs are summarized in Chapter 3 (3.5).

4.3.1 miRNAs as prognostic signatures for OS and RFS

Case-control approach:

A total of 1,423 miRNAs were annotated, of which 126 were retained after filtering for read counts from normal and tumor tissues. Eighty DE miRNAs were identified with FC $\geq$ 2.0 and FDR $\leq$ 0.05 (Appendix Table 9.1). As explained in the methods, these 80 miRNAs were treated as continuous variables and were subjected to univariate Cox analysis, followed by permutation test. Four miRNAs were associated with OS and two miRNAs were associated with RFS with permutation $p \leq 0.1$. The four and two miRNAs identified for OS (Table 4.1) and RFS (Table 4.2), respectively were used for constructing the risk score. A risk score cut-off point of 1.07 for OS was used to dichotomize the cases into low- ($\leq$ 1.07) and high-risk groups ($>$ 1.07). Similarly,

samples were grouped into the two risk groups based on the cut-off point estimated for RFS (0.72). Risk score was then treated as a categorical variable and entered into the univariate Cox model. Tumor stage, grade, age at diagnosis and TNBC status were considered as other clinical covariates and were first tested for their significance in the univariate Cox model. Tumor stage, grade and age at diagnosis were considered as potential confounders, and, irrespective of their significance in the univariate analysis, they were entered into the multivariate model along with the risk score. The higher-risk group was found to have both shorter OS (Hazard ratio, HR = 2.71, p = 0.004; Table 4.3, Figure 4.1) and RFS (HR = 2.27, p = 0.003; Table 4.3, Figure 4.2), after adjusting for confounders (tumor stage and age at diagnosis for OS and tumor stage for RFS).

Table 4.1 List of miRNAs significant for Overall Survival

miRNA ID	Univariate Cox p-value	Permuted p-value
hsa-miR-210-3p	0.01	0.02
hsa-miR-15a-5p	0.02	0.03
hsa-miR-660-5p	0.03	0.04
hsa-miR-146b-5p	0.04	0.05
hsa-miR-374a-3p	0.04	0.05
hsa-miR-374a-5p	0.04	0.06
hsa-miR-27a-3p	0.06	0.07
hsa-miR-574-3p	0.08	0.07
hsa-miR-221-3p	0.07	0.08
hsa-miR-196a-5p	0.07	0.09
hsa-miR-425-5p	0.05	0.10

Table 4.2 List of miRNAs significant for Recurrence Free Survival

miRNA ID	Univariate Cox p-value	Permuted p-value
hsa-miR-210-3p	0.01	0.02
hsa-miR-425-5p	0.05	0.08
hsa-miR-193b-3p	0.09	0.09
hsa-miR-15a-5p	0.08	0.10

Table 4: The tables above indicate the list of miRNAs associated with OS (Table 4.1) and RFS (Table 4.2) from both CC and CO approaches, along with their corresponding univariate Cox p-values and permuted p-values. All the miRNAs identified in CC approach were identified in the CO approach as well and are indicated in red.

Figure 4.1 miRNAs Kaplan Meier plot for Overall Survival (Case-control)

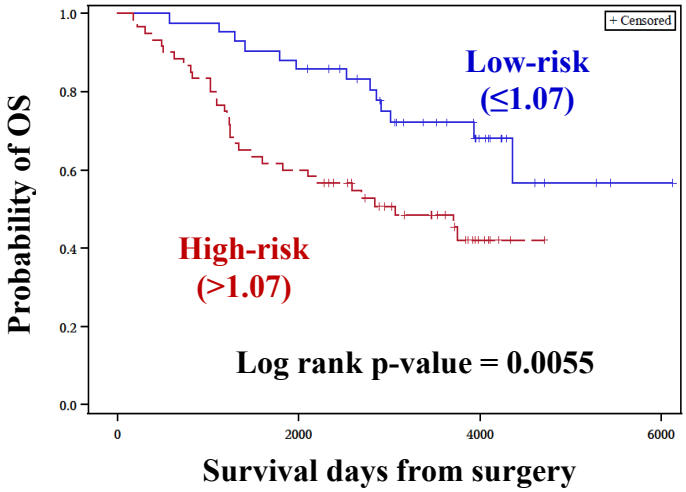


Figure 4.2 miRNAs Kaplan Meier plot for Recurrence Free Survival (Case-control)

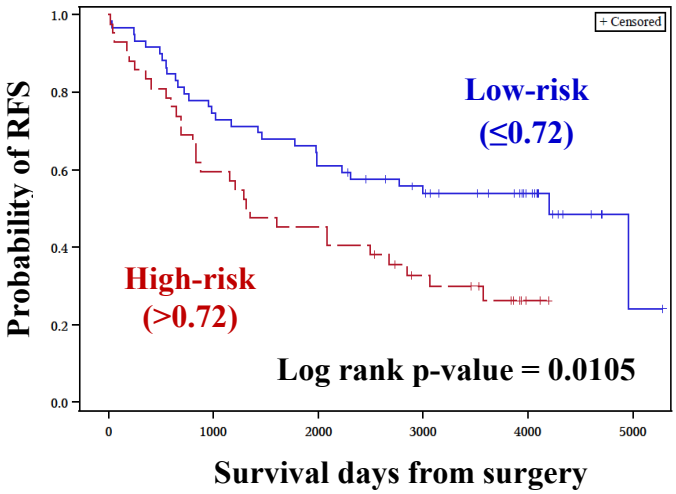


Figure 4: Kaplan-Meier plots for the CC approach using the risk score were constructed to determine the survival differences between low–risk and high–risk groups. Significant survival differences existed between the two risk groups, as indicated by the log–rank p–values. In both OS (Figure 4.1) and RFS (Figure 4.2), patients belonging to high–risk group showed poor prognoses.

Table 4.3 Univariate and Multivariate results of miRNAs from Case-control approach

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.44 (1.28 – 4.68)	0.01	2.71 (1.38 – 5.35)	0.004	1.95 (1.16 – 3.29)	0.01	2.27 (1.33 -3.88)	0.003
Tumor stage	0.42 (0.22 – 0.81)	0.01	0.36 (0.18 – 0.74)	0.01	0.42 (0.23 – 0.76)	0.01	0.34 (0.18 – 0.65)	0.001
Tumor grade	1.93 (0.99 – 3.75)	0.05			1.52 (0.88 – 2.63)	0.14		
Age at diagnosis	1.05 (1.02 – 1.09)	0.003	1.04 (1.01 – 1.07)	0.02	1.02 (0.99 – 1.05)	0.29		
TNBC status	0.88 (0.43 – 1.77)	0.71			0.75 (0.39 – 1.41)	0.37		

HR = Hazard Ratio; CI = Confidence Interval; TNBC = Triple Negative Breast Cancer

Table 4.3: miRNAs significant for OS (left panel) and RFS (right panel), identified from CC approach were used to construct a risk score. Receiver Operating Characteristics Curve was used to dichotomize samples into low and high-risk groups. Univariate Cox proportional hazards regression model was run for risk score and for other clinical parameters. In the multivariate analysis, risk score was significant with $p<0.05$ after adjusting for confounders. Multivariate analysis results indicate that patients belonging to high-risk group were at higher risk for death and recurrence (Hazard ratio > 1.0).

Case-only approach:

One hundred and forty seven miRNAs retained after filtering for read counts. These miRNAs were treated as continuous variables and were subjected to univariate Cox analysis followed by the permutation test. In this analysis, 11 miRNAs and 4 miRNAs were associated with OS (Table 4.1) and RFS (Table 4.2) respectively, and were used for constructing the risk score. A risk score cut-off point of 4.65 for OS was used to dichotomize the cases into low- ($\leq$ 4.65) and high-risk groups (> 4.65). Similarly, samples were grouped into two risk groups, based on the cut-off point estimated for RFS (1.17).

Risk score was then treated as a categorical variable and entered into the univariate Cox model. Similar to the case-control approach, the higher-risk group was found to have both shorter OS (HR = 2.76, p = 0.002; Table 4.4, Figure 4.3) and RFS (HR = 1.85, p = 0.02; Table 4.4, Figure 4.4), after adjusting for confounders (tumor stage for OS and RFS).

Figure 4.3 miRNAs Kaplan-Meier plot for Overall Survival (Case-only)

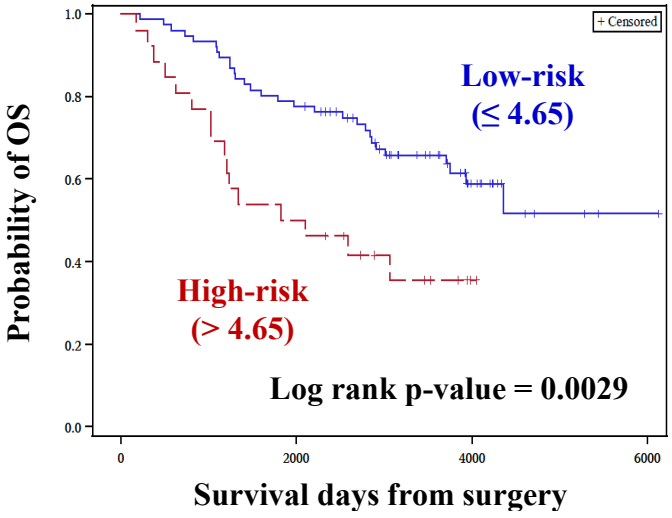
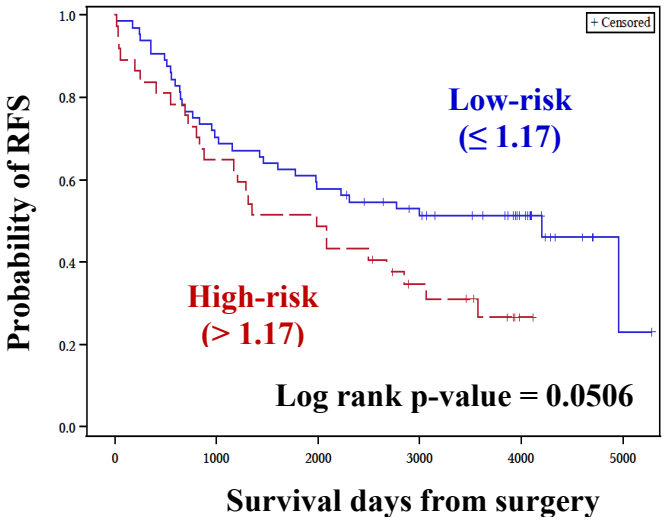


Figure 4.4 miRNAs Kaplan-Meier plot for Recurrence Free Survival (Case-only)



Figures 4.3 and 4.4: Kaplan-Meier plots for the Case-only approach using the risk score were constructed to determine the survival differences between low–risk and high–risk groups. Significant survival differences existed between the two risk groups, as indicated by the log–rank p–values. In both OS (Figure 4.3) and RFS (Figure 4.4), patients belonging to high–risk group showed poor prognoses.

Table 4.4 Univariate and Multivariate results of miRNAs from Case-only approach

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.48 (1.34 – 4.61)	0.004	2.76 (1.47 – 5.19)	0.002	1.68 (0.99 – 2.82)	0.05	1.85 (1.09 – 3.14)	0.02
Tumor stage	0.42 (0.22 – 0.81)	0.01	0.37 (0.19 – 0.72)	0.004	0.42 (0.23 – 0.79)	0.01	0.38 (0.20 – 0.71)	0.003
Tumor grade	1.93 (0.99 – 3.75)	0.05			1.52 (0.88 – 2.63)	0.14		
Age at diagnosis	1.05 (1.02 – 1.09)	0.003			1.02 (0.99 – 1.05)	0.29		
TNBC status	0.88 (0.43 – 1.77)	0.71			0.75 (0.39 – 1.41)	0.37		

HR = Hazard Ratio; CI = Confidence Interval; TNBC = Triple Negative Breast Cancer

Table 4.4: miRNAs significant for OS (left panel) and RFS (right panel), identified from CO approach were used to construct a risk score. Receiver Operating Characteristics Curve was used to dichotomize samples into low and high-risk groups. Univariate Cox proportional hazards regression model was run for risk score and for other clinical parameters. In the multivariate analysis, risk score was significant with $p<0.05$ after adjusting for confounders. Multivariate analysis results indicate that patients belonging to high-risk group were at higher risk for death and recurrence (Hazard ratio > 1.0).

4.3.2 Validation of OS-associated miRNAs in an external (TCGA) dataset

Eleven miRNAs that were significant for OS in the CO approach were validated using an external dataset (TCGA). Risk score was constructed using the eleven miRNAs. An optimal cut-off point was determined using ROC, to group samples into low (≤ -1.13) and high risk (> -1.13). Risk score which was considered as a categorical variable was significant with a p-value of 0.1 after adjusting for tumor stage. Similar to the discovery set, high risk group had shorter survival period with a HR of 2.07 (Figure 4.5, Table 4.5).

Table 4.5 Univariate and Multivariate results of miRNAs for Overall Survival (External Validation cohort/TCGA)

Parameter	Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.16 (0.92 – 5.05)	0.08	2.07 (0.87 – 4.92)	0.101
Tumor stage	0.32 (0.13 – 0.78)	0.01	0.26 (0.1 – 0.67)	0.005
Age at diagnosis	1.03 (1.003 – 1.06)	0.03		
TNBC status	0.63 (0.19 – 2.12)	0.46		

HR = Hazard Ratio; CI = Confidence Interval; TNBC = Triple Negative Breast Cancer

Table 4.5: Risk score was constructed in the TCGA dataset using the 11 miRNAs associated with OS and an optimal cut-off point was estimated using ROC, which dichotomized the samples into low and high-risk groups. Univariate Cox proportional hazards regression model was run for risk score and for other clinical parameters. In the multivariate analysis, risk score was significant with p = 0.1 after adjusting for tumor stage. Multivariate analysis results indicate that patients belonging to high-risk group were at higher risk for death (Hazard ratio > 1.0).

Figure 4.5 miRNAs Kaplan-Meier plot of the external dataset (TCGA)

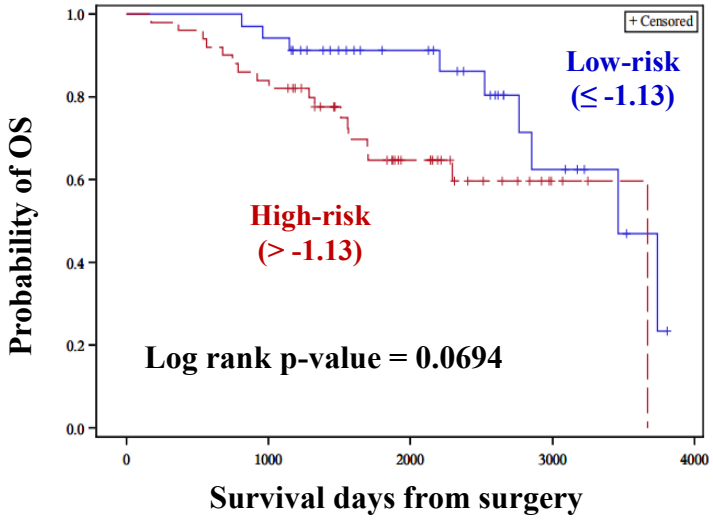


Figure 4.5: Kaplan-Meier plots were used to estimate OS in Case-only approach. Log rank test was performed to assess differences in survival between the two risk groups. Patients belonging to the high-risk group had shorter OS.

4.3.3 qRT-PCR validations of miR-99b-5p, miR-574-3p, miR-769-5p and miR-660-5p

The expressions of miR-99b-5p with a FC of -2.3, miR-574-3p with a FC of -5.8, miR-769-5p with a FC of -1.3 (down-regulated) and miR-660-5p with a FC of 12.8 (up-regulated) were tested in qRT-PCR to confirm the direction of effect and relative quantification agreement between NGS and qRT-PCR. Except for miR-660-5p, that was up-regulated (Figure 4.6), remaining three miRNAs were found to be significantly down-regulated in tumor tissues relative to normal samples in qRT-PCR experiments (Figure 4.7), which supported the NGS findings.

Figure 4.6 qRT-PCR validation of miR-660-5p

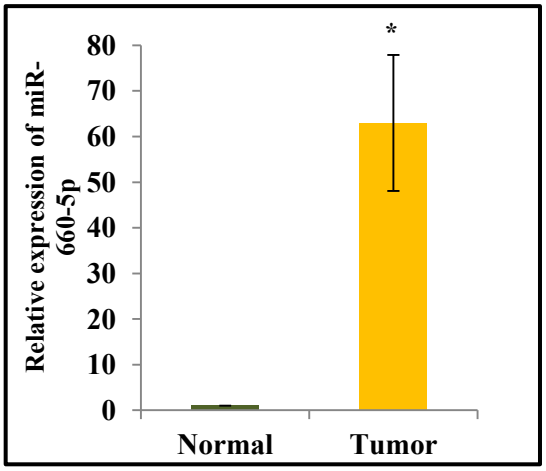
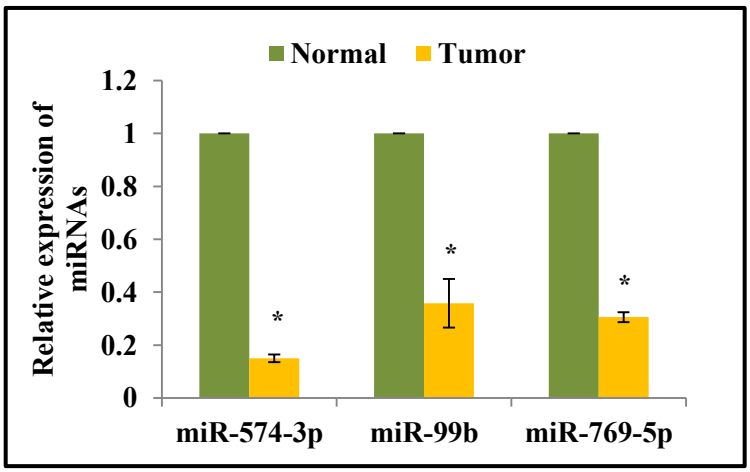


Figure 4.7 qRT-PCR validation of down-regulated miRNAs



Figures 4.6 and 4.7: One up-regulated miRNA (Figure 4.6. miR-660-5p, FC = 12.8) was validated in a subset of samples (9 normal samples and 56 tumor samples). Three down-regulated miRNAs (Figure 4.7. miR-574-3p, miR-99b-5p and miR-769-5p) were validated in a subset of samples (11 normal samples and 60 tumor samples). All the miRNAs were significantly (*=$p<0.05$) differentially expressed, similar to the results obtained in NGS platform. miR-574-3p and miR-660-5p were also found to be associated with Overall Survival.

4.3.4 Identification of potential targets for miRNAs and their role in cancer biology

The in-house transcriptome (mRNA) datasets available for BC were accessed (GEO accession ID GSE22820) [31] and analyzed for DE of mRNAs from two experiments, as outlined in the methods (refer 4.2.2). In the first experiment that included all the tumor samples, 5,399 genes (mRNAs) were DE, with 600 genes showing up-regulation and 4,799 genes showing down-regulation. In the experiment that included only matched tumor samples, 2,869 genes (mRNAs) were DE, of which 628 were up-regulated and 2,241 were down-regulated.

A combined total of 4,762 targets were predicted by TargetScan for the 12 miRNAs associated with OS and/or RFS. 1,038 targets (~22% of *in silico* predicted targets) overlapped with the DE genes identified from the experiment including all tumor samples, while in the matched dataset only 698 targets (~15% of in silico predicted targets) overlapped with the mRNA expression dataset. This low percent overlap between *in silico* and *in situ* comparisons is expected when breast tissue specific expression signatures filtered for histological and molecular subtypes are used to interrogate the potential interactions between miRNA-mRNA. The profiled interactions with transcriptome data also serve as an approach for functional validation of the miRNA targets within breast tissues and minimize the number of false positive targets identified.

In the experiment with all tumor samples, a total of 191 clusters were found and when filtered for clusters with enrichment score (ES) $\geq$ 1.3, a total of 75 clusters were found for the 12 miRNAs. However, in the matched dataset, a total of 168 clusters were found and when interrogated for gene ontology (GO) classifications with an ES $\geq$ 1.3, 57

clusters were retained. I have summarized the number of genes and gene ontology clusters identified from both the experiments as a comparison table in Table 4.6. I also compared the gene ontology terms for the targets of each of the miRNAs. Overall, I observed excellent concordance between the two experiments, suggesting that the use of matched or unmatched samples may not have a profound impact on the identification of gene targets for the miRNAs.

Table 4.6 Comparison of gene targets between all tumor samples and matched sample dataset

miRNA ID	# in silico targets	# targets overlapping between in silico prediction and in-house dataset			# Enrichment clusters		# clusters with Enrichment score ≥ 1.3	
		All tumor samples	Matched samples	Gene Overlap	All tumor samples	Matched samples	All tumor samples	Matched samples
miR-660-5p	149	31	25	20	8	2	5	1
miR-574-3p	13	3	2	2	0	0	0	0
miR-425-5p	212	30	24	17	11	2	2	0
miR-374a-3p; miR-374a-5p	680	123	110	90	30	22	11	9
miR-27a-3p	1212	193	183	141	51	47	19	13
miR-221-3p	446	64	60	45	19	25	11	13
miR-210-3p	32	8	6	6	1	1	1	0
miR-196a-5p	295	48	46	34	13	13	6	5
miR-193b-3p	222	34	30	26	7	4	1	0
miR-15a-5p	1275	213	181	148	48	48	19	15
miR-146b-5p	226	36	31	24	3	4	0	1

Table 4.6: Targets predicted for the 12 miRNAs were overlapped with gene expression dataset and compared using all the tumor samples and using only matched tumor sample datasets. Excellent concordance was observed in terms of number of gene targets and gene ontology clusters. Since only two targets were identified for miR-574-3p, gene ontology classification was not possible.

However, to maintain brevity and clarity, I am only summarizing the gene ontology classification table (Table 4.7) obtained for matched samples. From the clusters, statistically significant GO terms ($p < 0.05$) related to cancer were identified. Specifically, the following terms were interrogated: transcription, blood vessel development, angiogenesis, cell growth, cell morphogenesis, cell motion, cell migration, cell signaling, mammary gland development, cell differentiation, cell proliferation, cell division and cytoskeletal organization. Targets of 8 out of 12 miRNAs (miR-15a-5p, miR-27a-3p, miR-374a-3p, miR-374a-5p, miR-221-3p, miR-196a-5p, miR-146b-5p and miR-660-5p) were enriched for any one of the above-mentioned terms. Targets of miR-574-3p, miR-425-5p, miR-210-3p and miR-193b-3p were clustered with an ES ≤ 1.3 when matched miRNA-mRNA data sets were used and were therefore not probed further.

Table 4.7 Gene ontology classification of targets obtained from matched dataset

Gene ontology term	miRNA ID	Gene targets
Transcription	hsa-miR-660-5p	EGR2, NPAS3, ZBTB34, RFX4, EPAS1, NFAT5, ETV1, NR3C1, MEIS1
	hsa-miR-374a-3p; hsa-miR-374a-5p	CEBPA, HLF, ZBTB34, EGR2, BACH2, RFX4, EPAS1, ARID5B, ONECUT2, TLE4, CREB5, NFIX, NEUROG2, NR3C1, TCEAL7, LMX1A, MEIS1, EBF3, PNRC1, GBX2, HOXA10, SHOX, NFIB
	hsa-miR-221-3p	CDKN1C, SOX10, FOS, NTF3, GATA4, HOXA7, NFAT5, IGF1, GLI2, FOXP2, ZFP36, ZFP36L2, EIF4E3, QKI, TNRC6B
	hsa-miR-196a-5p	ING5, ERG, ZNF516, HOXA5, E2F7, BCL11A, HOXA7, HOXA9, TGFBR3, HABP4, IGF1, HMGA2, LIN28B, FOXP2
	hsa-miR-27a-3p	MEF2C, ING5, ZNF516, ZBTB34, E2F7, PPARG, ONECUT2, SOX7, EHF, PRDM16, SOX8, NPAS3, HOXA5, NR1D2, BCL11A, NFAT5, HOXA10, ERG, SMAD9, RUNX1T1, PPARGC1B, NRIP2, FOXP2, FOXN4, ZFHX4, HOXC11, ATF3, EBF3, BCORL1, NEUROD4, CAND1, NFIB
Cell morphogenesis	hsa-miR-374a-3p; hsa-miR-374a-5p	BMP2, EGR2, NTF3, ONECUT2, NEUROG2, L1CAM, LMX1A, SLIT3, SEMA5A, EPHA4, DMD, GBX2, CNTN4
	hsa-miR-221-3p	NTF3, PVRL1, DCX, GLI2, CXCL12
	hsa-miR-27a-3p	EGFR, SEMA6A, MAP1B, PRICKLE2, ONECUT2, LIFR, TGFBR3, RELN, NRXN1, NGFR, DCX, CACNA1A
Cell motion	hsa-miR-374a-3p; hsa-miR-374a-5p	SEMA5A, EPHA4, EGR2, NTF3, ARID5B, GBX2, NEUROG2, L1CAM, CNTN4, LMX1A, PPAP2B, SLIT3
	hsa-miR-221-3p	NTF3, PVRL1, WASF2, EMX2, IGF1, KIT, DCX, GLI2, CXCL12
	hsa-miR-196a-5p	PDGFRA, TGFBR3, IGF1, SEMA3A
	hsa-miR-27a-3p	RET, MET, IGF1, NRXN1, COL5A1, SEMA6A, BTG1, TGFBR3, NEUROD4, RELN, NGFR, DCX, PPAP2B
	hsa-miR-15a-5p	BDNF, PVRL1, PODXL, TGFBR3, IGF1, RELN, SEMA3A, LAMC1, CX3CL1, PPAP2A, FGF2, PPAP2B
Angiogenesis	hsa-miR-374a-3p; hsa-miR-374a-5p	SEMA5A, EPAS1, FGF9, GBX2, TGFA
	hsa-miR-15a-5p	RTN4, MEOX2, FGF9, PLCD1, FGF1, FIGF, FGF2
Cell migration	hsa-miR-221-3p	PDGFA, IGF1, KIT, CXCL12, PIK3R1
	hsa-miR-27a-3p	RET, BTG1, MET, TGFBR3, NEUROD4, RELN, DCX, PPAP2B, COL5A1
	hsa-miR-15a-5p	PODXL, TGFBR3, RELN, LAMC1, CX3CL1, PPAP2A, FGF2, PPAP2B

Table 4.7 (continued) Gene ontology classification of targets obtained from matched dataset

Gene ontology term	miRNA ID	Gene targets
Cell proliferation	hsa-miR-221-3p	ZFP36L2, PDGFA, EMX2, IGF1, KIT, NRG1, GLI2, CXCL12
	hsa-miR-196a-5p	TGFBR3, IGF1, FOXP2
	hsa-miR-15a-5p	TXNIP, FGFR1, FGF7, FGF9, E2F7, IGF1, FOXP2, PTHLH, BDNF, TRIM35, TBRG1, TGFBR3, ADAMTS1, RARB, LAMC1, AXIN2, PPAP2A, FGF1, NRG1, FIGF, FGF2, HTR2A
Cell signaling	hsa-miR-221-3p	NTF3, PVRL1, PDGFA, FGF14, GATA4, KCNA1, CACNB4, CXCL11, GLI2, NOVA1
	hsa-miR-27a-3p	STX1A, FGF14, CACNB2, NRXN1, GRIA4, LEP, SPRY2, ECE2, PDE7B, WISP1, HOXC11, FGF1, CACNA1A, NOVA1, DTNA
	hsa-miR-15a-5p	STX1A, KCNC4, FGF9, NLGN1, PTHLH, BDNF, WISP1, KIF1B, HOXC11, PVRL1, GRM7, FGF1, CHRNE, FGF2, HTR2A
Blood vessel development	hsa-miR-221-3p	RECK, PDGFA, WASF2, QKI, CXCL12
	hsa-miR-15a-5p	RTN4, RECK, MEOX2, FGF9, TGFBR3, QKI, PLCD1, FGF1, FIGF, FGF2, PPAP2B
Cytoskeleton organization	hsa-miR-146b-5p	PRC1, WASF3, TLN2, WASF2, ABL2
Response to estrogen stimulus	hsa-miR-27a-3p	PPARG, MAP1B, PDGFRA, MMP13, CCNA2
Positive regulation of cell differentiation	hsa-miR-27a-3p	LEP, LPL, ACVR2A, SMAD9, BTG1, CSF1, PPARG, MAP1B, NGFR
Regulation of cell division	hsa-miR-15a-5p	FGF7, FGF9, FGF1, FIGF, FGF2
Regulation of cell growth	hsa-miR-15a-5p	RTN4, EXTL3, WISP1, TSPYL2, SEMA3A, NRG1, FGF2, CRIM1
Mammary gland development	hsa-miR-221-3p	IGF1, NRG1, GLI2
Mesenchymal cell differentiation	hsa-miR-374a-3p; hsa-miR-374a-5p	BMP2, GBX2, CYP26A1

Table 4.7: The identified miRNAs significant for OS and RFS (n=12) from both the approaches were interrogated for mRNA targets, followed by identification of Gene ontology terms.

4.4 Discussion

In this study, I identified two miRNAs (miR-574-3p and miR-660-5p) as potential novel prognostic markers for BC, associated with OS. They have not been reported earlier for BC, for their association with either OS or RFS. Overall, from both the

approaches (CC and CO) adopted for the study, eleven miRNAs and four miRNAs were significant for OS and RFS, respectively. Out of the four miRNAs identified for RFS, three miRNAs (miR-210-3p, miR-425-5p and miR-15a-5p) were also significant for OS.

Although it is common to see either of the two approaches, i.e., either CC or CO methods for identifying prognostic markers, I have adopted both the approaches to identify the most suitable method for the study. As expected, higher numbers of miRNAs were identified as significant in the CO approach. Eleven miRNAs were significant for OS and four miRNAs were significant for RFS in the CO approach as opposed to four and two miRNAs significant for OS and RFS, respectively, in the CC method. miR-210-3p, miR-425-5p and miR-15a-5p were significant for both OS and RFS. A total of 12 non-redundant miRNAs were found to play a role in BC prognosis.

Overall, the differential expression in normal vs. tumor tissues and direction of effects show excellent agreement with what is known from published literature, as detailed below.

Novel prognostic miRNAs for BC

Of the 12 miRNAs identified in this study, two miRNAs (miR-574-3p and miR-660-5p) are potential novel prognostic markers for BC. Both the miRNAs were DE in a tumor vs. normal comparison, with miR-574-3p being down-regulated (FC = -5.8) and miR-660-5p being up-regulated (FC = 12.8) in the tumor samples. A similar direction of effect has been observed for miR-574-3p and miR-660-5p for ovarian cancer [32], colorectal cancer [33] and gastric cancer [34]; and chronic lymphocytic leukemia [35], respectively. However, this is the first report of a potential prognostic role for these

miRNAs in BC, although mechanistic insights are required to understand their contribution to tumorigenesis.

miRNAs with dual roles as tumor suppressor and oncogene

In this study, miR-15a-5p was found to be up-regulated in breast tissues (FC = 12.16) and the same direction of expression was observed in Kaposi sarcoma [36] and papillary thyroid carcinomas [37]. However, in other cancer types such as colorectal cancer [38], non-small-cell lung cancer (NSCLC) [39] and pituitary tumors [40], it is expressed in the opposite direction (down-regulation). Amongst BC reports, Kodahl et al. have reported an up-regulation of this miRNA [41] and a recent report by Shinden et al. has shown miR-15a as an independent prognostic marker for BC [42]. Similarly, miR-27a-3p, which was found to be up-regulated in tumors (FC = 6.45) in our study, is in accordance with the direction of expression observed in pancreatic cancer [43] and glioma [44]. Tang et al. have also reported miR-27a to be an oncomiR, the high expression of which promotes breast tumor growth and metastasis and is associated with poor OS in BC patients [45]. However, it is down-regulated in bladder cancer, compared with the normal samples [46]. The observations on miR-15a-5p and miR-27a-3p point to the dual roles of an oncogene and a tumor suppressor and their relative role may be governed in a tissue-specific manner.

miRNAs as oncogenes

I observed high expression (FC = 1.98) of miR-425-5p in breast tumors compared to the normal samples, which is concordant with the results published by Kodahl et al. for BC [41]. Likewise, Peng et al. have also observed the oncogenic function of miR-425,

which promotes cell proliferation, cell cycle progression, migration and invasion in gastric cancer [47].

Up-regulation of miR-146b in tumors and its adverse effect on survival has been demonstrated in lung cancer [48,49], thyroid carcinoma [50] and prostate cancer [51], among other cancer types. Interestingly, miR-146b-5p has also been reported to be up-regulated in BC, which is in accordance with my results (FC = 1.42) and is known to repress BRCA1 expression, thereby promoting cell proliferation [52].

miR-221 is a widely studied oncogene whose high expression is invariably associated with poor outcomes in several cancer types [53-55], including BC [56]. I also report the same direction of expression in tumor tissues with a FC of 1.27.

Cell proliferation, migration, invasion and metastasis have been found to be promoted in BC [57-59], glioblastoma [60,61], head and neck cancer [62] and gastric cancer [63,64] due to high expressions of miR-210, miR-196a and miR-374a (including miR-374a-3p and -5p), demonstrating their oncogenic potential. Their role as prognostic markers has also been studied in the above-mentioned cancer types. I was able to identify their prognostic significance following the CO approach, and these findings could have been missed if only the CC approach had been used. The read counts of the two groups (normal and tumor) revealed that these miRNAs were indeed present in higher amounts in tumors relative to the normal samples; the average read counts of miR-210-3p, miR-196a-5p, miR-374a-3p and miR-374a-5p in the normal samples were 2.5, 9.2, 0.7 and 1.09 respectively as against 59.7, 307.6, 46.1 and 108.9 for the tumor group. The lower read counts in normal samples have limited our ability to consider them in a CC study

due to our stringent filtering criteria. Overall, the patterns of DE and prognostic significance for the above miRNAs mirror observations from other cancer types.

miRNAs as tumor suppressors

In this study, apart from miR-574-3p, miR-193b-3p was also found to be down-regulated (FC = -4.3) in tumors compared to normal samples, which is in agreement with the studies on endometrioid adenocarcinoma [65], pancreatic cancer [66], oesophageal cancer [67] and gastric cancer [64]. Even in BC, Li et al. have reported a down-regulation of miR-193b in BC cell lines, and the low expression of miR-193b was found to be associated with shorter disease-free survival [68].

Functional roles of the identified prognostic miRNAs

The prognostic significance for recurrence or survival of an associated miRNA is better appreciated from the aspect of potential functional impact on cellular signaling and metabolic pathways, as these contribute to cell death, invasion and overall outcomes for the patient. Apart from functional insights, the potential for development of therapeutics is also important. Keeping these factors in mind, the following discussion is focused on the delineation of pathways using GO terms that are specifically enriched by the identified prognostic miRNAs.

Databases such as TargetScan, miRanda (http://www.microrna.org/) and PicTar (http://pictar.mdc-berlin.de/) have predicted mRNA targets, but a validation of the predicted targets adds more credence to *in silico* predictions. To this end, I first predicted the targets for all 12 miRNAs using the commonly used database - TargetScan; these were then compared with DE mRNAs obtained from the in-house BC transcriptome

dataset. GO terms were identified with a specific focus on terms pertaining to hallmarks of cancer. Interestingly, targets of eight miRNAs were found to be relevant for cell growth and development, indicating that these miRNAs may play key roles in tumorigenesis. Two targets (DAB2IP and SAMD4A) were found for miR-574-3p, of which DAB2IP is involved in apoptosis [69], cell survival [70], among other functions and SAMD4A functions as a translational regulator [71].

Validation of the identified signatures

In a biomarker study, a validation of the findings across different platforms is critical to rule out technical artifacts. Four miRNAs exhibiting different FC (lowest FC being -1.3) were validated using qRT-PCR, with two of the representative miRNAs identified as significant in survival analysis. The validation of representative miRNAs confirms cross-platform concordance and the relative utility of the signatures identified. However, validations using independent cohorts are also crucial for a biomarker study as they facilitate inter-study concordance of expression trends and signatures. NGS data for BC with a larger sample size and complete clinical information are limited in the public domain. I used the available data from TCGA project and applied stringent filtering criteria to obtain a dataset that would be comparable to the discovery set. A total of eleven miRNAs which were found to be associated with OS from the CO approach were considered for validation using the TCGA dataset. Multivariate analysis revealed that the risk score was significant with $p = 0.1$ after adjusting for tumor stage. Although for the initial analysis using discovery set I considered $p < 0.05$ as nominal, the TCGA dataset did not meet this threshold, presumably due to modest sample size (n=84) and events (n=27) compared to the discovery set (sample size, n=104 and events, n=46).

Nevertheless, I still observed the same direction of effect (Hazard Ratio), i.e., patients belonging to the high-risk group were associated with shorter survival period and this validates the initial observations from the discovery set.

Several differences existed between the discovery and validation datasets: (i) the NGS platform for discovery set was Genome Analyzer IIx where as for the validation set was HiSeq; (ii) the risk score cut-off point were estimated individually due to NGS platform differences; (iii) TCGA samples considered for this study were fresh frozen breast cancer tissues whereas the discovery set of breast cancer tissues were from FFPE blocks, (iv) information on tumor grade was not available for TCGA samples and (v) percent cellularity differences were also noted between the discovery and validation cohorts (see methods). However, despite these differences and other characteristics (Table 2.1), same direction of effects (Hazard Ratio) was observed in both the discovery and validation cohorts. The apparent lack of statistical significance (defined nominal value of 0.05) in the OS analysis attempted with TCGA data may be due to the differences between the two cohorts as well as to the limited sample size and limited number of events in the validation set affecting the power. Further validation of findings is warranted using independent cohorts and higher sample size and events. Overall, two novel miRNAs are reported as potential prognostic markers for breast cancer. Remaining miRNAs reported in this study showed excellent concordance to the published reports for their role in BC prognosis.

4.5 Conclusions

In summary, a total of twelve non-redundant miRNAs were identified to be associated with OS and/or RFS. As explained above, ten of the identified miRNAs have

been reported in literature as associated with BC prognosis and lends support to the findings in this independent study. However, two miRNAs (miR-574-3p and miR-660-5p) have not been reported previously for BC prognosis. The use of NGS platform to profile miRNAs on a whole genome level in BC has been limited thus far in literature and the data provided complements such efforts towards a comprehensive search for biomarkers. The miRNAs reported for OS have also been validated in independent dataset (TCGA) and functional characterization may help to understand the complex interplay of miRNA mediated gene regulation.

Overall, despite the increasing feasibility of profiling miRNAs and their role in prognostication, mechanistic insights in to the role of miRNAs, establishing gold standard approaches for analysis, and confirmation of these findings by independent laboratories within the context of confounding variables (histological and molecular heterogeneity, stage, grade and treatment) are needed to advance these promising biomarkers into clinical validation.

There is also a growing body of evidence that other small non-coding RNAs such as tRNAs [72], snoRNAs [73] and piRNAs [74] may contribute to tumorigenesis; however their role in BC prognosis is an area of active investigation. Therefore, a deeper exploration of their roles may pave the way for a comprehensive understanding of the small non-coding RNA classes, aiding in the discovery of newer diagnostic and prognostic biomarkers for BC.

miRNAs therefore served as a logical starting point for my thesis in that validation of the prognostic miRNAs identified to-date not only strengthens the study premise but is

also a reflection on the optimization of data mining approaches, statistical rigor, overall study design for use of NGS data for understanding the contribution of other sncRNAs to BC.

4.6 References

1. GLOBOCAN 2012 v 1.0, cancer incidence and mortality worldwide: IARC CancerBase no. 11. Journal of Endourology 2015;29:495-497.

2. Bertucci F, Birnbaum D. Reasons for breast cancer heterogeneity. J Biol 2008;7:6-6.

3. Ribelles N, Perez-Villa L, Jerez JM, Pajares B, Vicioso L, Jimenez B, et al. Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. Breast Cancer Res 2013;15:R98-R98.

4. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-2826.

5. Chang HY, Dimitry SAN, Sneddon JB, Hastie T, Tibshirani R, Sørlie T, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738-3743.

6. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.

7. Sotiriou C, Author, Wirapati P, Author, Loi S, Author, Harris A, Author, Fox S, Author, Smeds J, Author, et al. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-272.

8. Naderi A, Teschendorff AE, Barbosa-Morais N, Pinder SE, Green AR, Powe DG, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26:1507-1516.

9. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care? Nature Reviews Cancer 2007;7:545-553.

10. Bertoli G, Cava C, Castiglioni I. MicroRNAs: New biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. Theranostics 2015;5:1122-1143.

11. Bartel DP. Review: MicroRNAs. genomics, biogenesis, mechanism, and function. Cell 2004;116:281-297.

12. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 2004;116:843-854.

13. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 2004;116:855-862.

14. Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: MicroRNAs can up-regulate translation. Science 2007;318:1931-1934.

15. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37:495-500.

16. Rabinowits G, Gerçel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal MicroRNA: A diagnostic marker for lung cancer. Clin lung cancer 2009;10:42-46.

17. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008;18:997-1006.

18. Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

19. Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-microRNA signature. Gut 2010;59:579-585.

20. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 2013;19:4477-4487.

21. Cuk K, Zucknick M, Heil Jö, Madhavan D, Schott S, Turchinovich A, et al. Circulating microRNAs in plasma as early detection markers for breast cancer. Int J Cancer 2013;132:1602-1612.

22. Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-1184.

23. Kleivi Sahlberg K, Bottai G, Naume B, Burwinkel B, Calin GA, Børresen-Dale A, et al. A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. Clin Cancer Res 2015;21:1207-1214.

24. Tam S, de Borja R, Tsao M, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest 2014;94:350-358.

25. Liu J, Jennings SF, Tong W, Hong H. Next generation sequencing for profiling expression of miRNAs: Technical progress and applications in drug development. Journal of Biomedical Science and Engineering (JBiSE) 2011;4:666-676.

26. Farazi TA, Horlings HM, Ten Hoeve J,J., Mihailovic A, Halfwerk H, Morozov P, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res 2011;71:4443-4453.

27. Volinia S, Croce CM. Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. Proc Natl Acad Sci U S A 2013;110:7413-7417.

28. Liu N, Cui R, Sun Y, Guo R, Mao Y, Tang L, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. International Journal of Cancer 2014;134:1359-1368.

29. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 2001;25:402-408.

30. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009;4:44-57.

31. Bastani M, Vos L, Asgarian N, Deschenes J, Graham K, Mackey J, et al. A machine learned classifier that uses gene expression data to accurately predict estrogen receptor status. PLoS one 2013;8:1-9.

32. Wang L, Zhu M, Ren A, Wu H, Han W, Tan R, et al. A ten-MicroRNA signature identified from a genome-wide MicroRNA expression profiling in human epithelial ovarian cancer. PLoS one 2014;9:1-8.

33. Scarpati GDV, Calura E, Di Marino M, Romualdi C, Beltrame L, Malapelle U, et al. Analysis of differential miRNA expression in primary tumor and stroma of colorectal cancer patients. BioMed Research International 2014:1-8.

34. Su Y, Ni Z, Wang G, Cui J, Wei C, Wang J, et al. Aberrant expression of microRNAs in gastric cancer and biological significance of miR-574-3p. Int Immunopharmacol 2012;13:468-475.

35. Zhu D, Zhu W, Fang C, Fan L, Zou Z, Wang Y, et al. miR-181a/b significantly enhances drug sensitivity in chronic lymphocytic leukemia cells via targeting multiple anti-apoptosis genes. Carcinogenesis 2012;33:1294-1301.

36. Wu X, Pu X, Zhao Z, Zhao Y, Kang X, Wu W, et al. The expression profiles of microRNAs in kaposi's sarcoma. Tumor Biol 2015;36:437-446.

37. Liu X, He M, Hou Y, Liang B, Zhao L, Ma S, et al. Expression profiles of microRNAs and their target genes in papillary thyroid carcinoma. Oncol Rep 2013;29:1415-1420.

38. Xiao G, Tang H, Wei W, Li J, Ji L, Ge J. Aberrant expression of MicroRNA-15a and MicroRNA-16 synergistically associates with tumor progression and prognosis in patients with colorectal cancer. Gastroenterology Research & Practice 2014:1-8.

39. Tafsiri E, Darbouy M, Shadmehr MB, Zagryazhskaya A, Alizadeh J, Karimipoor M. Expression of miRNAs in non-small-cell lung carcinomas and their association with clinicopathological features. Tumor Biol 2015;36:1603-1612.

40. Renjie W, Haiqian L. MiR-132, miR-15a and miR-16 synergistically inhibit pituitary tumor cell proliferation, invasion and migration by targeting Sox5. Cancer Lett 2015;356:568-578.

41. Kodahl AR, Lyng MB, Binder H, Cold S, Gravgaard K, Knoop AS, et al. Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: A case control study. Molecular Oncology 2014;8:874-883.

42. Shinden Y, Akiyoshi S, Ueo H, Nambara S, Saito T, Komatsu H, et al. Diminished expression of MiR-15a is an independent prognostic marker for breast cancer cases. Anticancer Res 2015;35:123-127.

43. Wang W, Liu L, Li G, Chen Y, Li C, Jin D, et al. Combined serum CA19-9 and miR-27a-3p in peripheral blood mononuclear cells to diagnose pancreatic cancer. Cancer Prev Res (Phila) 2013;6:331-338.

44. Xu W, Liu M, Peng X, Zhou P, Zhou J, Xu K, et al. miR-24-3p and miR-27a-3p promote cell proliferation in glioma cells via cooperative regulation of MXI1. Int J Oncol 2013;42:757-766.

45. Tang W, Zhu J, Su S, Wu W, Liu Q, Su F, et al. MiR-27 as a prognostic marker for breast cancer progression and patient survival. PLoS ONE 2012;7:1-7.

46. Jiang X, Du L, Wang L, Li J, Liu Y, Zheng G, et al. Serum microRNA expression signatures identified from genome-wide microRNA profiling serve as novel noninvasive biomarkers for diagnosis and recurrence of bladder cancer. Int J Cancer 2015;136:854-862.

47. Wei-Zhao Peng, Ma R, Wang F, Yu J, Zhi-Bin Liu. Role of miR-191/425 cluster in tumorigenesis and diagnosis of gastric cancer. International Journal of Molecular Sciences 2014;15:4031-4048.

48. Rani S, Gately K, Crown J, O'Byrne K, O'Driscoll L. Global analysis of serum microRNAs as potential biomarkers for lung adenocarcinoma. Cancer Biol Ther 2013;14:1104-1112.

49. Raponi M, Dossey L, Jatkoe T, Wu X, Chen G, Fan H, et al. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. Cancer Res 2009;69:5776-5783.

50. Lee YS, Lim YS, Lee J, Wang S, Park H, Kim SY, et al. Differential expression levels of plasma-derived miR-146b and miR-155 in papillary thyroid cancer. Oral Oncol 2015;51:77-83.

51. Walter BA, Valera VA, Pinto PA, Merino MJ. Comprehensive microRNA profiling of prostate cancer. J Cancer 2013;4:350-357.

52. Garcia AI, Buisson M, Bertrand P, Rimokh R, Rouleau E, Lopez BS, et al. Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. EMBO Mol Med 2011;3:279-290.

53. Kim B, Hong SW, Kim A, Choi SH, Yoon SO. Prognostic implications for high expression of oncogenic microRNAs in advanced gastric carcinoma. J Surg Oncol 2013;107:505-510.

54. Li J, Wang Y, Yu W, Chen J, Luo J. Expression of serum miR-221 in human hepatocellular carcinoma and its prognostic significance. Biochem Biophys Res Commun 2011;406:70-73.

55. Teixeira AL, Ferreira M, Silva J, Gomes M, Dias F, Santos JI, et al. Higher circulating expression levels of miR-221 associated with poor overall survival in renal cell carcinoma patients. Tumor Biol 2014;35:4057-4066.

56. Falkenberg N, Anastasov N, Rappl K, Braselmann H, Auer G, Walch A, et al. MiR-221/-222 differentiate prognostic groups in advanced breast cancers and influence cell invasion. Br J Cancer 2013;109:2714-2723.

57. Cai J, Guan H, Fang L, Yang Y, Zhu X, Yuan J, et al. MicroRNA-374a activates wnt/ß-catenin signaling to promote breast cancer metastasis. J Clin Invest 2013;123:566-579.

58. Hong L, Yang J, Han Y, Lu Q, Cao J, Syed L. High expression of miR-210 predicts poor survival in patients with breast cancer: A meta-analysis. Gene 2012;507:135-138.

59. Rothé F, Ignatiadis M, Chaboteaux C, Haibe-Kains B, Kheddoumi N, Majjaj S, et al. Global MicroRNA expression profiling identifies MiR-210 associated with tumor proliferation, invasion and poor clinical outcome in breast cancer. PLoS ONE 2011;6:1-13.

60. Yan W, Li R, Liu Y, Yang P, Wang Z, Zhang C, et al. MicroRNA expression patterns in the malignant progression of gliomas and a 5-microRNA signature for prognosis. Oncotarget 2014;5:12908-12915.

61. Yang G, Han D, Chen X, Zhang D, Wang L, Shi C, et al. MiR-196a exerts its oncogenic effect in glioblastoma multiforme by inhibition of IκBα both in vitro and in vivo. Neuro Oncol 2014;16:652-661.

62. Suh Y, Raulf N, Gaken J, Lawler K, Urbano TG, Bullenkamp J, et al. MicroRNA-196a promotes an oncogenic effect in head and neck cancer cells by suppressing annexin A1 and enhancing radioresistance. International Journal of Cancer 2015;137:1021-1034.

63. Xu X, Wang W, Su N, Zhu X, Yao J, Gao W, et al. miR-374a promotes cell proliferation, migration and invasion by targeting SRCIN1 in gastric cancer. FEBS Lett 2015;589:407-413.

64. Mu Y, Tang S, Sun W, Gao W, Wang M, Su X. Association of miR-193b down-regulation and miR-196a up-regulation with clinicopathological features and prognosis in gastric cancer. Asian Pac J Cancer Prev 2014;15:8893-8900.

65. Wu W, Lin Z, Zhuang Z, Liang X. Expression profile of mammalian microRNAs in endometrioid adenocarcinoma. Eur J Cancer Prev 2009;18:50-55.

66. Li J, Kong F, Wu K, Song K, He J, Sun W. miR-193b directly targets STMN1 and uPA genes and suppresses tumor growth and metastasis in pancreatic cancer. Molecular Medicine Reports 2014;10:2613-2620.

67. Hummel R, Sie C, Watson DI, Wang T, Ansar A, Michael MZ, et al. MicroRNA signatures in chemotherapy resistant esophageal cancer cell lines. World J Gastroenterol 2014;20:14904-14912.

68. Li X-, Yan P-, Shao Z-. Downregulation of miR-193b contributes to enhance urokinase-type plasminogen activator (uPA) expression and tumor progression and invasion in human breast cancer. Oncogene 2009;28:3937-3948.

69. Yu L, Tumati V, Tseng S, Hsu F, Kim DN, Hong D, et al. DAB2IP regulates autophagy in prostate cancer in response to combined treatment of radiation and a DNA-PKcs inhibitor. Neoplasia 2012;14:1203-1212.

70. Shen Y, Kong Z, Wan F, Wang H, Bian X, Gan H, et al. Downregulation of DAB2IP results in cell proliferation and invasion and contributes to unfavorable outcomes in bladder cancer. Cancer Science 2014;105:704-712.

71. de Haro M, Al-Ramahi I, Jones KR, Holth JK, Timchenko LT, Botas J. Smaug/SAMD4A restores translational activity of CUGBP1 and suppresses CUG-induced myopathy. PLoS Genetics 2013;9:1-10.

72. Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. RNA 2013;19:461-6.

73. Chen L, Han L, Wei J, Zhang K, Shi Z, Duan R, et al. SNORD76, a box C/D snoRNA, acts as a tumor suppressor in glioblastoma. Sci Rep 2015;5:8588-8588.

74. Yan H, Wu Q, Sun C, Ai L, Deng J, Zhang L, et al. piRNA-823 contributes to tumorigenesis by regulating de novo DNA methylation and angiogenesis in multiple myeloma. Leukemia (08876924) 2015;29:196-206.

5 Piwi-interacting RNAs and PIWI genes as novel prognostic markers for Breast cancer

5.1 Introduction

Piwi-interacting RNAs (piRNAs, 24 – 32 nt in length) belong to a class of small regulatory RNAs that include microRNAs (miRNAs) and small interfering RNAs (siRNAs) [1]. Mature forms of these RNAs associate with biogenesis pathway proteins such as Argonaute (AGO) class of proteins: miRNAs and siRNAs with AGO proteins [2] and piRNAs with PIWI proteins [3-6] to guide target specific gene regulation [7,8]. Gene regulation exerts control at transcriptional and post-transcriptional levels and piRNAs, in association with PIWI proteins, are involved in both levels [9,10]. For a long time, the only roles of PIWI proteins were believed to be in the regulation of transposons [11] and in the maintenance and development of germinal stem cells [12] ; however, the functions of piRNAs and PIWI proteins as epigenetic regulators have recently started to emerge. It is now known that PIWI proteins, which are guided by piRNAs bind to specific targets (based on sequence specific complementarity) and recruit chromatin modifiers to enable transcriptional repression [13]. Apart from this, a direct association between the piRNA–PIWI protein complex and stem cell development and maintenance has been established [14]. Cancer stem cells form a critical fraction of a tumor mass, are required for incessant cell proliferation, and may underlie resistance to drugs and radiation; accordingly, cancer stem cells are believed to contribute to tumor recurrence [15,16].

 I would like to thank Jennifer Dufour for assistance in providing access to specimens and assay protocols and Mahalakshmi Kumaran for training in the use of miRanda algorithm.

The role of the piRNA–PIWI protein complex in post-transcriptional gene regulation is also slowly garnering attention. Although the exact mechanism remains elusive, investigators initially have reported the sequence specific complementary binding of a piRNA to a target messenger RNA (mRNA) at the 3' untranslated region (UTR) and subsequent gene regulation, in a manner similar to that of miRNAs [17-19]. It is increasingly being recognized that the sequence based complementarity may not be restricted to 3' UTR and may expand to 5'UTR, the coding sequence or even the introns [20]. Given the diverse functions of piRNAs and PIWI proteins, it is evident that these molecules may also contribute to tumorigenesis [10].

Human homologues of PIWI proteins (originally described as P-element induced wimpy testis in Drosophila) identified thus far are PIWIL1 (HIWI), PIWIL2 (HILI), PIWIL3 and PIWIL4 (HIWI2) [21]. Although the expression of PIWI proteins in somatic tissues has been known since 1998, our major understanding of these molecules stem from germ cells. Only recently, have researchers demonstrated their possible involvement in tumorigenesis. Aberrant expressions of these genes and proteins in malignancy have been associated with hallmarks of cancer and have also shown promise as potential prognostic and diagnostic markers for different cancer types [22]. In this regard, the differential expression of piRNAs and therefore their oncogenic or tumor suppressor roles have also been observed in various cancer types [19,20], and a few studies have highlighted their association with clinicopathological factors [23]. An even smaller number of studies have reported the relevance of piRNAs as prognostic/diagnostic markers [24-26]; however, the study designs of the majority of these studies are limited to candidate piRNA molecules or are challenged with limited sample sizes.

Given the current knowledge that piRNAs and PIWI genes (i) are abundantly expressed in somatic tissues, (ii) are potential biomarkers for cancer and (iii) are involved in gene regulation and in normal developmental processes, extensive profiling and characterization studies are needed to understand the contribution of these molecules to tumorigenesis. The contribution of both piRNAs and PIWI genes to breast cancer has not been comprehensively studied and is the focus of this chapter. I hypothesized that varying levels of piRNAs and their upstream biogenesis pathway (PIWI) genes contribute to breast tumorigenesis and act as prognostic markers for breast cancer. The specific objectives were (i) to identify differentially expressed piRNAs and PIWI gene transcripts (mRNAs) (hereafter referred to as PIWI genes) in breast tumor tissues relative to normal (reduction mammoplasty) breast tissues, (ii) to identify piRNAs and PIWI genes as prognostic markers (outcomes: overall survival, OS and recurrence free survival, RFS) and (iii) to identify complementary gene (mRNA) targets at the 3' UTR for the piRNAs associated with breast cancer prognosis.

5.2 Materials and methods

Details on clinical characteristics of samples used for the study (discovery and external validation cohorts), isolation of total RNA, profiling of piRNAs and statistical analysis involved in identifying piRNAs with prognostic significance and the methods employed for external validation of the identified prognostic markers are explained in detail in chapter 2 (sections 2.2, 2.3, 2.4, 2.5 and 2.6). Methods that are more specific for this chapter and those that have not been explained elsewhere are explained below:

5.2.1 Validation of piRNA

qRT-PCR experiments were performed in collaboration with Dr. Kovalchuk's laboratory in University of Lethbridge and I analyzed and interpreted the data.

One randomly chosen representative piRNA showing prognostic significance (hsa_piR_009051) was validated with the total RNA isolated from normal and tumor samples. qRT-PCR was performed using an iScript Select cDNA Synthesis Kit (Bio-Rad) and a SsoFast EvaGreen Supermix (Bio-Rad) according to manufacturers' instructions. Primer for the selected piRNA was designed as described elsewhere [27] and the sequence is as follows: piR009051-F: 5'-GCA GAG TGT AGC TTA ACA CAA AG-3', piR-009051-R: 5'-CCA GTT TTT TTT TTT TAG TTG GGT-3'. RNU6-2 served as loading control and the primer sequences are RNU6-2-F: 5'-CGC TTC GGC AGC ACA TAT AC-3', RNU6-2-R: 5'-AGG GGC CAT GCT AAT CTT CT-3'. All assays were done in triplicates, data was analyzed using the $2^{-\Delta\Delta Ct}$ method [28], and results are shown as fold induction of piRNA.

5.2.2 PIWI genes as prognostic markers for breast cancer

To identify PIWI genes with prognostic relevance, I accessed the in-house gene (mRNA) expression dataset generated using Agilent microarray platform for ten normal breast tissues (obtained from reduction mammoplasty) and 141 breast tumor tissues from gene expression omnibus (GSE22820) [29]. The data was quantile normalized and log2 transformed using PGS. Differential expression analysis was performed using one-way ANOVA to observe the expression patterns of the four human homologues of PIWI genes (PIWIL1 – PIWIL4). Survival analysis was performed for OS and RFS since there

were 42 deaths and 77 recurrence events in this dataset. Treating the four genes as continuous variables, univariate Cox regression analysis was carried out; PIWI genes with $p \leq 0.15$ were used for constructing a risk score and ROC estimated the optimal cut-off point for patient stratification into low and high-risk groups. Risk score was then treated as dichotomous variable; univariate and multivariate analysis was performed, considering tumor stage, grade, age at diagnosis and TNBC status as potential confounders.

5.2.3 Identification of gene targets for significant piRNAs and their functional roles

Of the eight prognostically significant piRNAs, six were DE and were of immediate interest for gene target prediction. Recent evidence has suggested (i) interaction between piRNAs and mRNAs through base-pair complementarity and (ii) a possible inverse correlation between piRNA expression and its corresponding mRNA targets [19,20]. Since all the six piRNAs (selected for target prediction) were up-regulated, I extracted only the down-regulated genes (mRNAs), with FC > 2.0 and FDR 0.05 (as determined by one-way ANOVA) from the in-house gene expression dataset. The breast tissues (tumor tissue and normal reduction mammoplasty specimens) used in both the NGS and mRNA expression experiments are from the same clinics in Alberta. Although other possible mechanisms of action viz., binding to coding exons and 5'UTR has been suggested [20], I focused initially on the putative binding of piRNAs to 3'UTR of coding genes. I extracted the fasta sequences of the 3'UTR of all the down-regulated genes from Ensembl database (GRCh37, http://grch37.ensembl.org/index.html) [30] and obtained the fasta sequences of the six piRNAs from piRNA Bank (hg 19, http://pirnabank.ibab.ac.in/)

[31]. As such, there are no target prediction databases available for piRNAs. However, predictions based on the list of input genes (in this study, down-regulated genes in breast cancer tissues) were obtained using miRanda v 3.3a algorithm [32], with alignment score ≥ 170 and energy threshold ≤ -20 kcal/mol [20]. These stringent cut-offs have been adopted from a previous study by Hashim et al., that has successfully predicted target complementary sequences for a given set of piRNAs using these cut-offs [20]. While alignment score is indicative of the degree of complementarity shared between piRNA and target mRNA, free energy is indicative of the stability of the RNA pair. Therefore higher alignment score and lower free energy value is important to identify potentially stable piRNA-mRNA pairs. Potential functional insights of the targets (with a focus on biological processes) identified were obtained using DAVID bioinformatics tool (http://david.abcc.ncifcrf.gov/) [33] and I report gene ontology (GO) terms related to cancer with $p < 0.05$ in the current study.

5.3 Results

Details on profiling of piRNAs and identification of DE piRNAs from small RNA sequencing data are summarized in chapter 3 (3.5)

5.3.1 piRNAs are potential independent prognostic markers for breast cancer

Case-control approach:

In summary, 676 piRNAs were profiled from breast tissues (normal and tumor tissues inclusive) and 42 were retained after filtering for read counts in the CC approach. 25 piRNAs were DE (Appendix Table 9.1), distributed as 17 up-regulated piRNAs and 8 down-regulated piRNAs with FC $\geq$ 2.0 and FDR $\leq$ 0.05.

Of the 25 DE piRNAs, three piRNAs each were significant (permuted p value ≤ 0.1) in the univariate analysis for OS (Table 5.1) and RFS (Table 5.2) and were used to construct the individual risk scores. Two piRNAs (i.e. hsa_piR_009051 and hsa_piR_021032) were significant for both OS and RFS. The receiver operating characteristics curve (ROC) estimated cut-off points for OS and RFS were 2.04 and 0.07, respectively, dichotomizing the patients into low–risk (≤ 2.04 for OS and ≤ 0.07 for RFS) and high–risk (> 2.04 for OS and > 0.07 for RFS) groups. The risk scores were found to be significant after adjusting for tumor stage and age at diagnosis for OS (Table 5.3) and tumor stage for RFS (Table 5.3). Patients belonging to the high–risk group were associated with poor OS (Figure 5.1) and RFS (Figure 5.2).

Table 5.1 List of piRNAs significant for Overall Survival

piRNA ID	Univariate Cox p-value	Permuted p-value
hsa_piR_009051	0.01	0.01
hsa_piR_021032	0.01	0.03
hsa_piR_015249	0.06	0.07
hsa_piR_020541	0.07	0.09

Table 5.2 List of piRNAs significant for Recurrence Free Survival

piRNA ID	Univariate Cox p-value	Permuted p-value
hsa_piR_017061	0.02	0.02
hsa_piR_009051	0.03	0.05
hsa_piR_021032	0.03	0.06
hsa_piR_004153	0.08	0.06
hsa_piR_017716	0.09	0.08
hsa_piR_019914	0.09	0.09

Tables 5.1 and 5.2: Overall, four piRNAs and six piRNAs were significant for OS (Table 5.1) and RFS (Table 5.2), respectively from the CO approach. However, the CO approach identified piRNAs also included all of the piRNAs significant in the CC approach (three for OS and three for RFS) and are indicated in red color.

Figure 5.1 piRNAs Kaplan-Meier plots for Overall Survival (Case-control)

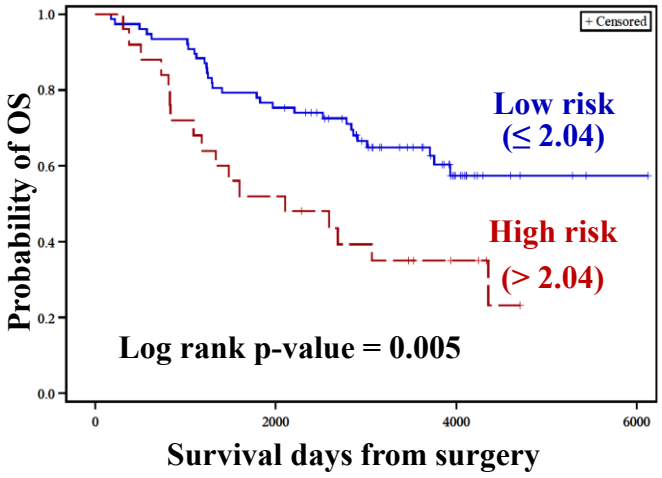
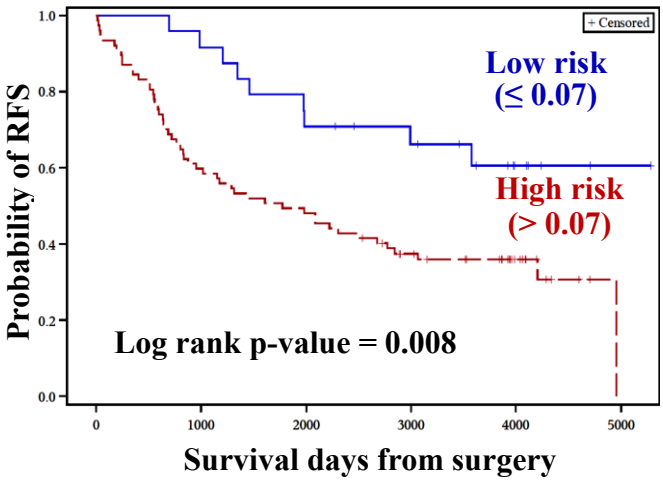


Figure 5.2 piRNAs Kaplan-Meier plots for Recurrence Free Survival (Case-control)



Figures 5.1 and 5.2: Risk scores were constructed using piRNAs significant in univariate Cox analysis with permuted p-value ≤ 0.1. Samples were dichotomized into low and high risk groups based on ROC estimation of optimal cut-off point (indicated in parenthesis). Patients belonging to high-risk group were associated with poor OS (Figure 5.1) and poor RFS (Figure 5.2), with log-rank p value < 0.05.

Table 5.3 Univariate and multivariate results of piRNAs from Case-control approach

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.31 (1.27 – 4.22)	0.01	2.29 (1.24 – 4.27)	0.01	2.53 (1.25 – 5.16)	0.01	2.79 (1.36 – 5.69)	0.005
Tumor stage	0.40 (0.21 – 0.78)	0.01	0.42 (0.21 – 0.84)	0.02	0.38 (0.20 – 0.71)	0.003	0.34 (0.18 – 0.63)	0.001
Tumor grade	2.01 (1.04 – 3.89)	0.04			1.58 (0.92 – 2.74)	0.1		
Age at diagnosis	1.06 (1.02 – 1.09)	0.001	1.04 (1.01 – 1.08)	0.01	1.02 (0.99 – 1.05)	0.21		
TNBC status	0.99 (1.16 – 3.29)	0.98			0.84 (0.45 – 1.55)	0.58		

HR = Hazards ratio; CI = Confidence interval

Table 5.3: Univariate and multivariate Cox analysis results for OS (left panel) and RFS (right panel) in case–control approach is represented. Patients belonging to high–risk group were associated with poor prognosis (HR > 1) and the risk score showed promise as potential independent prognostic factor ($p < 0.05$).

Case-only approach:

665 piRNAs were expressed with at least one read count in any one of the tumor samples and of these, 53 were retained with ≥ 10 read counts and expressed in at least 90% of the tumor samples. The raw data was adjusted for batch effects. Four and six piRNAs (from the 53 filtered piRNAs) were significant in the univariate analysis for OS (Table 5.1) and RFS (Table 5.2) with a permuted p-value ≤ 0.1. The risk scores were constructed using the four and six piRNAs for OS and RFS, respectively. ROC based estimation of the cut–off point further dichotomized the patients into two groups: low–risk (≤ 2.44 for OS and ≤ -0.54 for RFS) and high–risk (> 2.44 for OS and > -0.54 for

RFS). For both outcomes, (i) the risk score showed p–value significance in the univariate and multivariate analyses (Table 5.4) after adjusting for potential confounders (tumor grade and age at diagnosis for OS and tumor stage for RFS) and (ii) the high–risk group patients showed poor OS (Figure 5.3) and RFS (Figure 5.4).

Figure 5.3 piRNAs Kaplan-Meier plots for Overall Survival (Case-only)

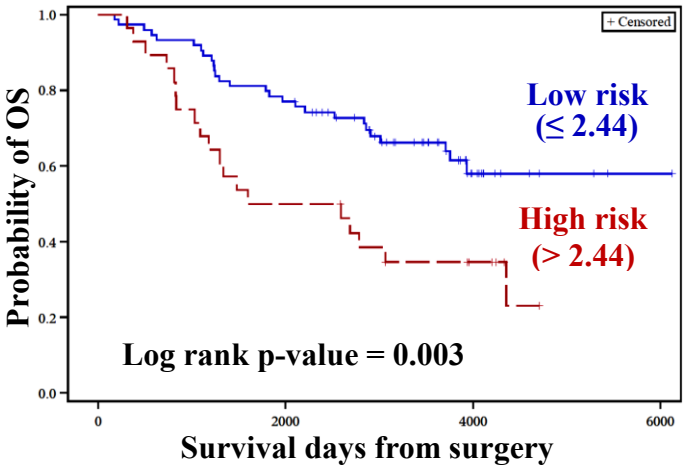
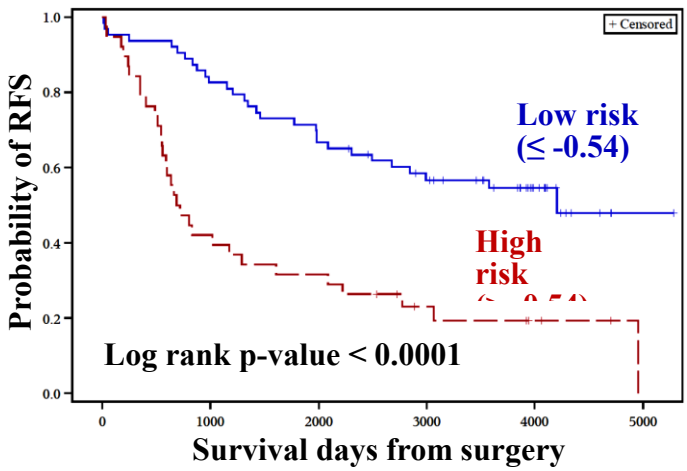


Figure 5.4 piRNAs Kaplan-Meier plots for Recurrence Free Survival (Case-only)



Figures 5.3 and 5.4: Risk scores were constructed using piRNAs significant in univariate Cox analysis with permuted p-value ≤ 0.1. Samples were dichotomized into low and high risk groups based on ROC estimation of optimal cut-off point (indicated in parenthesis). Patients belonging to high-risk group were associated with poor OS (Figure 5.3) and RFS (Figure 5.4), with log-rank p value < 0.05.

Table 5.4 Univariate and multivariate results of piRNAs for case–only approach

Parameter	Overall Survival				Recurrence free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.36 (1.31 – 4.26)	0.004	2.09 (1.15 – 3.79)	0.02	3.08 (1.84 – 5.16)	<0.0001	3.07 (1.84 – 5.14)	<0.0001
Tumor stage	0.40 (0.21 – 0.78)	0.01			0.38 (0.20 – 0.71)	0.003	0.39 (0.21 – 0.72)	0.003
Tumor grade	2.01 (1.04 – 3.89)	0.04	2.01 (1.03 – 3.92)	0.04	1.58 (0.92 – 2.74)	0.1		
Age at diagnosis	1.06 (1.02 – 1.09)	0.001	1.06 (1.02 – 1.09)	0.001	1.02 (0.99 – 1.05)	0.21		
TNBC status	0.99 (0.50 – 1.95)	0.98			0.84 (0.45 – 1.55)	0.58		

HR = Hazards ratio; CI = Confidence interval; TNBC = Triple Negative Breast Cancer

Table 5.4: Univariate and multivariate Cox analysis results for OS (left panel) and RFS (right panel) in case–only approach are represented. Patients belonging to high–risk group were associated with poor prognosis (HR > 1) and the risk score showed promise as potential independent prognostic factor ($p < 0.05$).

5.3.2 The risk score for OS was significant in the external validation dataset

I extracted the batch–adjusted normalized counts of the four piRNAs (significant for OS in the discovery cohort) from the 84 samples in The Cancer Genome Atlas (TCGA) dataset. A risk score was constructed for OS, and the ROC based estimation of the cut–off point dichotomized the samples into low–risk (≤ -0.18) and high–risk (> -0.18) groups. Similar to the results obtained in the discovery cohort, the risk score showed promise as a potential independent prognostic factor (Table 5.5), and the patients in the high–risk group were significantly associated with poor OS (Figure 5.5; $p<0.01$).

Figure 5.5 piRNAs Kaplan-Meier plot for external/TCGA dataset (Overall Survival)

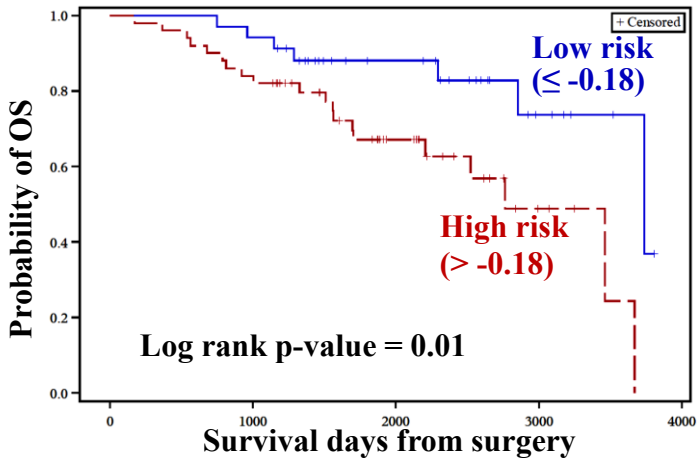


Figure 5.5: Risk score was constructed using the piRNAs significant for OS (identified from the discovery cohort). External/TCGA dataset also showed similar direction of effect with statistical significance, confirming the results obtained in the discovery cohort.

Table 5.5 Univariate and multivariate results of piRNAs for Overall Survival (External validation/TCGA dataset)

Parameter	Overall Survival			
	Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	3.02 (1.21 – 7.59)	0.02	3.22 (1.22 – 8.52)	0.02
Tumor stage	0.32 (0.13 – 0.78)	0.01	0.34 (0.14 – 0.88)	0.03
Age at diagnosis	1.03 (1.003 – 1.06)	0.03	1.04 (1.01 – 1.07)	0.006
TNBC status	0.63 (0.19 – 2.12)	0.46		

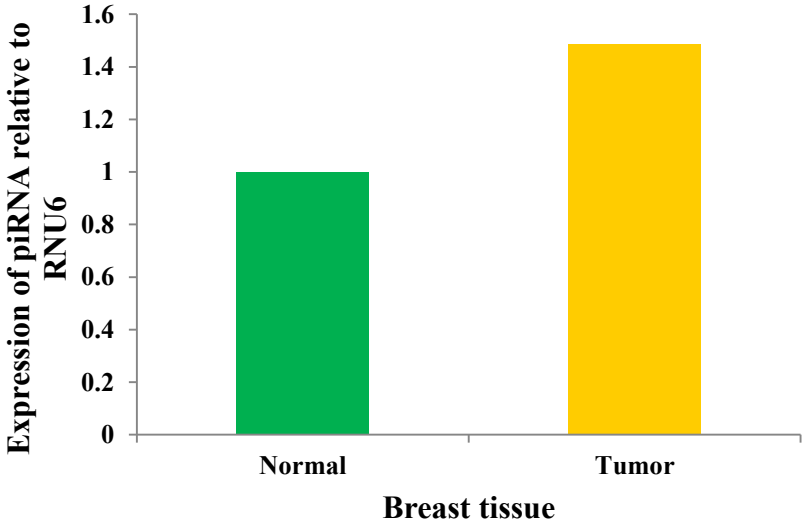
HR = Hazards ratio; CI = confidence interval; TNBC = Triple Negative Breast Cancer

Table 5.5: Risk score constructed using four piRNAs (identified as significant for OS in discovery cohort) was adjusted for tumor stage and age at diagnosis and was found to be significant with $p < 0.05$ in TCGA dataset (external validation set).

5.3.3 Concordance of hsa_piR_009051 expression between NGS and qRT-PCR

hsa_piR_009051 was found to be up-regulated in tumor tissues with a FC of 4.38. The direction of expression of this piRNA was confirmed in a subset of the samples by qRT-PCR (Figure 5.6) with a FC of 1.49 and p-value of 0.09, validating the findings from NGS. Although the obtained p-value was not less than 0.05 due to sample size limitations (isolated RNA from FFPE was available in limited quantities due to several experimental validations attempted for all the profiled sncRNAs), nevertheless this experiment confirmed the direction of expression of hsa_piR_009051.

Figure 5.6 qRT-PCR confirmation of hsa_piR_009051



5.3.4 PIWI genes are promising prognostic markers for breast cancer

All four human homologues of PIWI genes were expressed in the in-house breast cancer gene expression dataset. Comparison with normal breast tissues revealed that two genes (PIWIL1 and PIWIL3) were up-regulated and the remaining two (PIWIL2 and PIWIL4) were down-regulated in tumor tissues (Table 5.6). The up-regulated PIWI genes did not show statistical significance between normal and breast tumor tissues.

Nevertheless, the expression of PIWI genes in breast (somatic) tissues was confirmed. Since these proteins are involved in piRNA biogenesis, I hypothesized that an aberrant expression of these genes in breast cancer may contribute to abnormal expression of piRNAs. Since piRNAs showed prognostic relevance, I hypothesized that genes coding for PIWI proteins may also be involved in breast cancer prognosis. Of the four PIWI genes, only the PIWIL3 and PIWIL4 genes were significant in the univariate analysis for OS and were used to construct a risk score. Similar to the piRNA analysis, ROC was used to estimate the optimal cut–off point for dichotomization of patients into low–risk (≤ 0.56) and high–risk (> 0.56) groups. The risk score was significant for OS after adjusting for age at diagnosis and TNBC status (Table 5.7). In the case of RFS, PIWIL3 gene was found to be significant. The potential of PIWIL3 gene as an independent prognostic marker was confirmed in the multivariate analysis (Table 5.7). For both OS (Figure 5.7) and RFS (Figure 5.8), patients belonging to the high–risk group were found to have shorter survival.

Table 5.6 Differential expression of PIWI genes

PIWI gene	Fold change	Direction of expression	p-value
PIWIL1	1.56	Up-regulated in tumor	0.06
PIWIL2	-2.51	Down-regulated in tumor	6.97E-5
PIWIL3	1.44	Up-regulated in tumor	0.12
PIWIL4	-1.95	Down-regulated in tumor	0.0044

Table 5.6. Of the four human homologs of PIWI gene, PIWIL1 and PIWIL3 were up-regulated but were not statistically significant. PIWIL2 and PIWIL4 genes were down-regulated and were statistically significant with $p < 0.05$.

Figure 5.7 PIWI Kaplan-Meier plot for Overall Survival

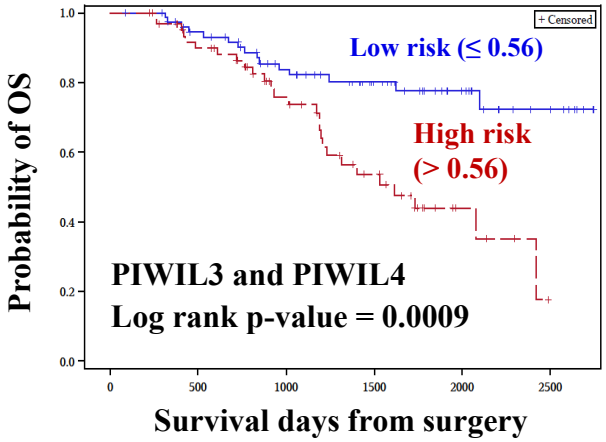
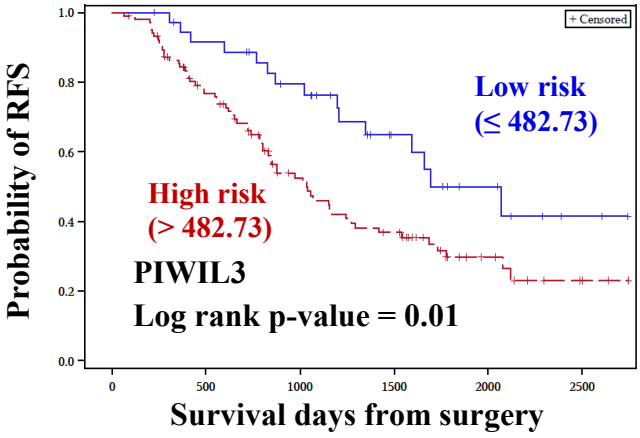


Figure 5.8 PIWI Kaplan-Meier plot for Recurrence Free Survival



Figures 5.7 and 5.8: PIWIL3 and PIWIL4 genes were significant for OS and were used for constructing a risk score, whereas PIWIL3 alone was significant for RFS. Patients were dichotomized into low and high-risk groups based on ROC estimated cut-off point. Patients belonging to high-risk group were associated with poor OS (Figure 5.6) and RFS (Figure 5.7).

Table 5.7 Univariate and multivariate results of PIWI genes

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score (for OS) PIWIL3 (for RFS)	2.82 (1.49 – 5.33)	0.002	2.19 (1.14 – 4.22)	0.02	2.07 (1.17 – 3.64)	0.01	2.09 (1.18 – 3.71)	0.01
Tumor stage	0.62 (0.24 – 1.57)	0.31			0.56 (0.28 – 1.11)	0.09		
Tumor grade	2.31 (1.1 – 4.83)	0.03			1.75 (1.06 – 2.9)	0.03		
Age at diagnosis	1.04 (1.02 – 1.07)	0.001	1.04 (1.02 – 1.07)	0.001	1.01 (0.99 – 1.03)	0.22		
TNBC status	3.33 (1.77 – 6.26)	0.0002	2.35 (1.15 – 4.79)	0.02	1.72 (1.07 – 2.79)	0.03		

HR = Hazards ratio, CI = Confidence interval; TNBC = Triple Negative Breast Cancer

Table 5.7: Univariate analysis was performed, considering PIWI genes as continuous variables. Two PIWI genes were significant for OS with $p \leq 0.15$ and were used for constructing a risk score, PIWIL3 alone was significant for RFS with $p \leq 0.15$. Risk score for OS and PIWIL3 for RFS were considered as categorical variables and were found to be significant in univariate and multivariate analyses using Cox proportional hazards regression model.

5.3.5 piRNAs repress gene expression

Recent evidence suggests that piRNAs, in a mechanism similar to miRNAs, may regulate gene expression through base pair complementarity, and as such, very few studies have identified the corresponding gene targets for specific piRNAs [19,20]. For this study, I only considered prognostically significant piRNAs (eight non–redundant piRNAs in total from OS and RFS) and focused on the inverse correlations between piRNAs and their targets. Of the eight piRNAs, only six were differentially expressed (all were more than 1.5 FC) and were of immediate interest for target predictions. Since all six were up-regulated in tumors, relative to normal tissues, I extracted the 3'UTR sequences of all the

down–regulated genes (n = 2,735) identified in our gene expression dataset. Using miRanda algorithm v3.3a and applying the cut–offs, a total of 350 (306 non–redundant) gene targets for six piRNAs were identified (Appendix Table 9.3). I did not consider matched samples (between the piRNA data and the mRNA data) alone for target prediction, but instead utilized all the samples from the gene expression dataset since the previous study on miRNA-mRNA target identifications using the same mRNA dataset did not reveal profound differences in the overall functional terms identified for the targets (Section 4.3.4, Table 4.6) [34]. The identified gene targets were enriched for angiogenesis, transcription, cell signaling, cytoskeleton organization, membrane transport and organization (Table 5.8).

Table 5.8 Identification of piRNA gene targets and their functional roles

piRNA ID	# of gene targets	# of GO clusters	Targets	GO term
hsa_piR_009051	10	1	SSBP2, FOXO4, NR5A2, ZNF177, ZNF765	Regulation of transcription
hsa_piR_021032	180	27	KCNMA1, CAV2, NRP1, SCN2B, GLRA3, AKAP9, NRXN1, ATP1A2, ESR2, PARK2, KCNMB1, SEMA5A, LEP, PDE7B, NPTX1, KIF1B, KCNN1, SLC22A3	Cell-cell signaling
			SEMA5A, NRP1, PLXDC1, LEPR, CCBE1, ROBO4, TNFSF12	Angiogenesis
			KCNMA1, TRPM3, TRPM6, CUBN, SLC16A12, ATP1A2, SLC26A4, SLC2A4, SLC22A3, SLC25A37, KCNH8, SV2B, NALCN, SLC25A26	Transmembrane transport
			TXNIP, KCNMA1, CAV2, GSTM3, LEPR	Response to estrogen stimulus
			TRIOBP, SHROOM4, MRAS, NEDD9, FGD5, ARHGAP26, FGD4	Actin cytoskeleton organization
			LEP, LEPR, GAB1, PDCD4, FGD4	Regulation of MAPKKK cascade
			GAB1, PDCD4, FGD4	Regulation of JUN kinase activity
hsa_piR_015249	1	0	FOXP2	Transcription
hsa_piR_004153	42	6	ALPL, CALCR, CAV1	Response to glucocorticoid stimulus
hsa_piR_017716	72	7	ALPL, PPARA, GNG2, FOXO4, ACVR1C, SLC34A2	Response to hormone stimulus
			LAMA4, EPAS1, TNFSF12, ANGPTL4	Blood vessel development
			KCNK17, SLC23A2, P2RX3, KCNMB1, SLC34A2, ATP13A4, GRID1	Ion transport
hsa_piR_019914	45	7	EREG, LEPR, PLCD3, CXCL12	Angiogenesis
			EREG, LEPR, IGF1, GHR	Positive regulation of signal transduction
			LY75, ARRB1, EHD2, GHR	Membrane organization

GO = Gene Ontology; GO clusters represent biological processes; GO term includes cancer related terms with $p < 0.05$

5.4 Discussion

Prognostic significance of eight piRNAs for breast cancer are reported for the first time. Four and six piRNAs were found to be associated with OS and RFS, respectively, among which two piRNAs were common for OS and RFS. I also successfully validated the prognostic significance of piRNAs associated with OS in an external dataset (TCGA). Gene targets for possible regulation by candidate piRNAs have also been identified. Although PIWI proteins have been studied by others as prognostic/diagnostic markers for other cancer types, their prognostic relevance in breast cancer has not been examined. This is the first study to demonstrate association of PIWI genes (as a proxy for PIWI proteins) with OS and RFS for breast cancer. Overall, this is the first study to comprehensively understand the significance of piRNAs and PIWI genes as prognostic markers for breast cancer using large and independent datasets with complete clinical annotation and a long follow–up period. In this study, I have successfully captured the pathway of events and individual entities up-stream and down-stream of the piRNA biogenesis.

The clinical relevance of piRNAs was first apparent when they were reported to be associated with clinicopathological factors such as lymph node status [23], and TNM stage [24]. Nonetheless, our understanding of their contribution as prognostic markers is rudimentary and warrants further exploration. In this study, eight piRNAs were identified as novel prognostic markers for breast cancer. To date, there has only been one study that has utilized sequencing data to interrogate piRNAs for breast cancer prognosis [26]. In this study by Martinez et al., piRNAs associated with OS were identified for eleven cancer types, including breast cancer. This study is therefore the first to identify piRNAs

associated with RFS as well as OS. I compared the eight prognostically significant piRNAs with their study findings and found that hsa_piR_009051 and hsa_piR_017061 were prognostically significant for renal clear cell carcinoma and colon adenocarcinoma, respectively. hsa_piR_021032 was significantly associated with renal clear cell carcinoma and lung squamous cell carcinoma prognoses. Significance of the remaining five piRNAs in cancer prognosis remains unknown till date.

An important observation from this study is that we may obtain a more holistic picture of piRNAs associated with outcomes if we adopt a case–only approach. Case-control approach focuses on identifying prognostic markers which are differentially expressed [35,36]. However, case-only approach interrogates the entire dataset in an unbiased manner [37-39] and may thus yield higher number of prognostic markers. I observed the same in this study, where, with the case–only method, four and six piRNAs were obtained for OS and RFS, respectively as opposed to three piRNAs each for OS and RFS. The piRNAs identified in the case–only approach included the ones identified from the case–control approach as well. Therefore, adopting a case–only approach may provide a more comprehensive understanding of the markers under investigation.

Another major finding of the study was the identification of genes coding for PIWI proteins as potential prognostic markers for breast cancer. Of the four human homologues of PIWI genes, two genes (PIWIL3 and PIWIL4) showed associations with OS, and PIWIL3 alone showed an association with RFS. Earlier studies reported the prognostic significance of PIWIL1 in soft-tissue sarcoma [40] and glioma [41]. High expression of PIWIL2 transcript was found to be associated with decreased survival rate in colorectal cancer [42] and has also been found to contribute to cisplatin resistance in ovarian cancer [43].

Reports on the clinical significance of PIWIL3 and PIWIL4 remain scarce, and in particular, this is the first study to identify the contribution of PIWIL3 and PIWIL4 genes to breast cancer prognosis. Further replication studies are warranted to better define their prognostic roles.

The functional importance of PIWI proteins and piRNAs is no longer restricted to the regulation of transposons or the maintenance and development stem and germ cells. For instance, PIWIL1 and PIWIL2 genes have been observed to promote cell proliferation in gastric [44] and breast tumors [45]. Similarly, piRNAs have also exhibited involvement in several key cellular mechanisms [19,46]. Based on previous studies that piRNAs inhibit gene expression, analogous to miRNAs, I identified 306 gene targets (and their roles) for six piRNAs using the in-house gene expression dataset (Table 5.8 and Appendix Table 9.3). From the functional classification, it may be inferred that the piRNAs actively contribute to tumorigenesis by regulating genes involved in several pathways contributing to the development of cancer. However, I did not restrict the analysis to gene ontology terms alone that identified terms related to cancer. I looked at the targets identified for every piRNA individually and found piRNA-mRNA pairs playing important roles in methylation, oxidative stress, and cell adhesion, among others, the deregulation of which may contribute to an imbalance in cellular homeostasis. An interesting observation was that hsa_piR_021032 was found to share complementary sequence with target PIWIL2 (a member of the human PIWI genes), showing alignment score > 170 and energy score < -20 kcal/mol (Appendix Table 9.3). This PIWI gene was observed to be down–regulated in the gene expression dataset and hsa_piR_021032 was found to be up-regulated in the tumor tissues, suggesting a possible repression of the

PIWI gene by the piRNA. This proposed mechanism of PIWI regulation by piRNAs is novel and requires further validation.

5.5 Conclusions

Using a cohort with complete clinical annotation and long–term follow–up, I identified piRNAs and PIWI genes as novel prognostic markers for breast cancer. Identifying piRNA gene targets from breast tissue datasets is rare in the literature, and this study may open up research on the characterization of these piRNA–mRNA pairs. Deregulation of piRNAs and the involvement of the identified targets in key cellular mechanisms suggest that piRNAs may be important contributors to breast tumorigenesis. This is also the first time that a possible regulatory mechanism of PIWI genes by piRNAs has been observed, but it remains to be established if this regulation is through direct interaction or a complex network. Biomarker studies on piRNAs and PIWI genes and proteins are promising fields of research. Since piRNAs have exhibited stability in body fluids such as blood [47], serum and plasma [48], they may also serve as effective circulating biomarkers. With improving profiling platforms, availability of clinical samples with extensive clinical annotations, will likely contribute to identification of additional piRNAs, furthering our understanding of their mechanistic and prognostic contributions to breast cancer and other diseases.

5.6 References

1. Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: Its biogenesis and functions. Annu Rev Biochem 2015;84:405-433.

2. Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. The argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev 2002;16:2733-2742.

3. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian piwi proteins. Nature 2006;442:199-202.

4. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 2006;442:203-207.

5. Grivna ST, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20:1709-1714.

6. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science 2006:313:363-367.

7. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van dE, et al. Article: A role for piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell 2007;129:69-82.

8. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nature Reviews Molecular Cell Biology 2009:126.

9. Pal-Bhadra M, Bhadra U, Birchler JA. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in drosophila. Mol Cell 2002;9:315-327.

10. Ng KW, Anderson C, Marshall EA, Minatel BC, Enfield KSS, Saprunoff HL, et al. Piwi-interacting RNAs in cancer: Emerging functions and clinical utility. Molecular Cancer 2016;15:1-13.

11. Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the drosophila melanogaster flamenco gene. Genetics 2004;166-1313-21.

12. Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev 1998;12:3715-3727.

13. Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, et al. Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev 2007;21:2300-2311.

14. Sharma AK, Nelson MC, Brandt JE, Wessman M, Mahmud N, Weller KP, et al. Human CD34(+) stem cells express the hiwi gene, a human homologue of the drosophila gene piwi. Blood 2001;97:426-434.

15. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nature Reviews Cancer 2005:275.

16. Merlos-Suarez A, Barriga FM, Jung P, Iglesias M, Cespedes MV, Rossell D, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell stem cell 2011;8:511-524.

17. Peng L, Song L, Liu C, Lv X, Li X, Jie J, et al. piR-55490 inhibits the growth of lung carcinoma by suppressing mTOR signaling. Tumour Biol 2015;37:2749-2756.

18. Watanabe T, Lin H. Posttranscriptional regulation of gene expression by piwi proteins and piRNAs. Mol Cell 2014;56:18-27.

19. Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, et al. Identification of novel piRNAs in bladder cancer. Cancer Lett 2015;356:561-567.

20. Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-9910.

21. Sasaki T, Shiohama A, Minoshima S, Shimizu N. Identification of eight members of the argonaute family in the human genome. Genomics 2003;82:323-330.

22. Suzuki R, Honda S, Kirino Y. PIWI expression and function in cancer. Front Genet 2012;3:204-204.

23. Huang G, Hu H, Xue X, Shen S, Gao E, Guo G, et al. Altered expression of piRNAs and their relation with clinicopathologic features of breast cancer. Clinical & translational oncology 2013;15:563-568.

24. Busch J, Ralla B, Jung M, Wotschofsky Z, Trujillo-Arribas E, Schwabe P, et al. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. Journal of Experimental & Clinical Cancer Research (17569966) 2015;34:1-11.

25. Martinez VD, Enfield KSS, Rowbotham DA, Lam WL. An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence. Gastric Cancer 2015;19:660-665.

26. Martinez VD, Vucic EA, Thu KL, Hubaux R, Enfield KSS, Pikor LA, et al. Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. Sci Rep 2015;5:10423-10423.

27. Busk PK. A tool for design of primers for microRNA-specific quantitative RT-qPCR. BMC Bioinformatics 2014;15:2-22.

28. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 2001;25:402-408.

29. Bastani M, Vos L, Asgarian N, Deschenes J, Graham K, Mackey J, et al. A machine learned classifier that uses gene expression data to accurately predict estrogen receptor status. PLoS ONE 2013;8:1-9.

30. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl genomes 2016: More genomes, more complexity. Nucleic Acids Res 2016;44:D574-D580.

31. Sai Lakshmi S, Agrawal S. piRNABank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173-D177.

32. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in drosophila. Genome Biol 2003;5:R1-R1.

33. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009;4:44-57.

34. Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics 2015;16:735-735.

35. Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-1184.

36. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 2013;19:4477-4487.

37. Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

38. Liu N, Cui R, Sun Y, Guo R, Mao Y, Tang L, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. International Journal of Cancer 2014:1359.

39. Su Y, Ni Z, Wang G, Cui J, Wei C, Wang J, et al. Aberrant expression of microRNAs in gastric cancer and biological significance of miR-574-3p. Int Immunopharmacol 2012;13:468-475.

40. Taubert H, Greither T, Kaushal D, Würl P, Bache M, Bartel F, et al. Expression of the stem cell self-renewal gene hiwi and risk of tumour-related death in patients with soft-tissue sarcoma. Oncogene 2007;26:1098-1100.

41. Sun GA, Wang YY, Sun LH, Luo H, Liu N, Fu Z, et al. Clinical significance of hiwi gene expression in gliomas. Brain Res 2011;1373:183-188.

42. Oh SJ, Kim SM, Kim YO, Chang HK. Clinicopathologic implications of PIWIL2 expression in colorectal cancer. Korean journal of pathology 2012;46(4):318-323.

43. Wang Q, Han C, Milum K, Wani AA. Stem cell protein Piwil2 modulates chromatin modifications upon cisplatin treatment. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 2011;708:59-68.

44. Liu XY, Sun Y, Guo JP, Ma H, Li JY, Dong B, et al. Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. International journal of cancer 2006;118:1922-1929.

45. Lee JH, Jung C, Javadian-Elyaderani P, Schweyer S, Schütte D, Shoukier M, et al. Pathways of proliferation and antiapoptosis driven in breast cancer stem cells by stem cell protein piwil2. Cancer Res 2010;70:4569-4579.

46. Cheng J, Guo J, Xiao B, Miao Y, Jiang Z, Zhou H, et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. Clinica Chimica Acta 2011;412:1621-1625.

47. Cui L, Lou Y, Zhang X, Zhou H, Deng H, Song H, et al. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. Clin Biochem 2011;44:1050-1057.

48. Yang X, Cheng YD, Lu Q, Wei JF, Yang HW, Gu M. Detection of stably expressed piRNAs in human blood. International journal of clinical and experimental medicine 2015;8:13353-13358.

6 Genome wide profiling of transfer RNAs and their role as novel prognostic markers for breast cancer

6.1 Introduction

The discovery that only 2% of the human genome encodes for proteins (the coding portion) and that the remaining 98% (the non-coding portion) harbor sequences with structural, regulatory and functional relevance, dispelled the long-held belief that these sequences should be considered as "junk DNA" [1]. Amongst the non-coding portion of the genome which gets transcribed but not translated, two major classes of RNA exist based on size: long non-coding RNAs (> 200 nt) and small non-coding RNAs (sncRNAs < 200 nt) [2]. Both the classes of RNA contribute to post-transcriptional level of gene regulation. Several subcategories of sncRNAs exist, including microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), piwi-interacting RNAs (piRNAs) small nuclear RNAs (snRNAs) and transfer RNAs (tRNAs) [3].

While much of the focus has been on miRNAs [4], functional significance of other RNAs is less explored in cellular processes and for their potential roles as prognostic markers in cancer. Transfer RNAs (tRNAs) are a 73-92 nt long class of sncRNAs [3] that play a crucial role in protein synthesis. A total of 625 tRNA genes have been identified so far in the human genome, of which 506 are tRNAs that decode standard amino acids, three are selenocysteine tRNAs, three are suppressor tRNAs, three are tRNAs with undetermined or unknown isotypes and 110 are tRNAs predicted to be pseudogenes [5].

Apart from playing a role in protein translation, recent discoveries have suggested that tRNAs may play a vital role in activation of protein kinase GCN2 [6], regulation of apoptosis [7], and protein degradation [8]. Furthermore, processing of the 3' or 5' ends of mature or precursor tRNAs have given rise to another class of small RNAs called tRNA derived fragments (tRFs) [9]. Previous studies have demonstrated that tRFs are not degradation by-products but are functional molecules that arise during stress conditions [10]. Relative variations in expression levels of tRFs in tumor cells as compared to normal cells [11], and their role in silencing gene expression, thereby influencing cell proliferation [9] or metastasis [12] implies that they may also contribute to tumorigenesis. Interestingly, there is also evidence indicating that tRF may possess characteristics of a miRNA, both structurally and functionally (by regulating gene expression) [13]. Similar to miRNAs, tRFs have also recently showed promise as prognostic marker for prostate cancer [14], thus expanding the repertoire of tRNA functions but their clinical relevance to BC remains unexplored. While miRNAs are known to interact with mRNAs directly and promote gene expression regulation [15,16], recent studies have demonstrated contributions of tRNAs to post-transcriptional gene expression regulation. For instance, Maute et al, have identified a functionally active tRNA derived microRNA (miRNA) that represses the expression of protein coding gene by means of sequence complementarity to mRNA [13]. tRNAs may also act as a source for another molecule called piwi-interacting RNA (piRNA) [17], which are equally considered as master regulators of gene expression as previous studies have assigned a similar role to miRNAs [18,19]. This further expands the functions of tRNAs, warranting the need for a deeper exploration into this class of sncRNA.

Dysregulation of protein synthesis machinery has been observed in several tumor cells and has been found to be one of the major contributors for malignant transformation of cells [20]. Specifically, over expression of RNA polymerase III and its products (including tRNAs) has been observed in breast and ovarian cancers [21-23]. Studies on the consequences of aberrant expression of tRNAs have demonstrated that over-expression of initiator tRNA can drive cell proliferation, resulting in oncogenic transformation [24]. As such, tRNAs are now recognized for their pivotal role in tumorigenesis, though a comprehensive understanding of their diverse roles in the biology of cancer is far from complete.

Despite their discovery in 1956 [25,26], not many studies have focused on the comprehensive profiling of tRNAs and explored their potential to serve as biomarkers for cancer. Pavon-Eternod et al. were the first to profile tRNAs using a microarray platform, to demonstrate that over expression of tRNAs is a hallmark of breast cancer (BC) and have postulated their potential utility as biomarkers for BC [27]. However their significance as prognostic markers for BC remains unexplored to date. In fact, the prognostic potential of tRNAs has not been investigated for any type of cancer. Although there has been a considerable progress in creating personalized treatment strategies for BC patients, based on their ER, PR or Her2 receptor expression status, a subset of patients continue to experience recurrence, leading to mortality. Factors contributing to inter-individual variations in response to treatments and eventual clinical outcomes (Overall Survival, OS; Recurrence Free Survival, RFS) may be ascribed in part, to the heterogeneous nature of breast cancer (in terms of histological and molecular subtypes and morphologies) [28,29]. The continuing discoveries of additional molecular subsets of BC (based on deep

sequencing of tumor genomes) has called for the identification of novel biomarkers or combinations of biomarkers that may perform better than the traditional markers alone, in terms of prognostication or prediction. These molecular signatures may guide the development of target therapies and in the selection of treatment.

In this study, I hypothesized that relative variation in expression levels of tRNAs contribute to inter-individual differences in disease trajectory and in eventual treatment outcomes. Small RNA libraries generated from 11 apparently healthy normal breast tissue samples (obtained from reduction mammoplasty surgery) and 104 breast tumor samples with complete clinical information [30] were sequenced. The specific objectives were to (i) profile and identify differentially expressed (DE) tRNAs, (ii) investigate the role of tRNAs as prognostic markers for BC treatment outcomes (OS and RFS), (iii) validate the signatures in an external dataset, and lastly (iv) investigate the contribution of tRNAs to gene regulation. I confirm that tRNAs are globally up-regulated in BC and report for the first time, the prognostic significance of 27 tRNAs.

6.2 Materials and Methods

I have summarized the clinical characteristics of samples (discovery cohort and external validation/TCGA cohort) used for the study (section 2.2), RNA isolation (section 2.3) and sequencing protocols (section 2.4), sequencing data analysis (section 2.5) and statistical analysis (section 2.6) involved in identifying and validating prognostic markers in Chapter 2.

I have explained below the methods specific for tRNAs and these have not been explained elsewhere in the thesis.

6.2.1 Cross platform concordance to validate expression of select tRNAs

qRT-PCR experiments were performed in collaboration with Dr. Kovalchuk's laboratory in University of Lethbridge and I analyzed and interpreted the data.

Total RNA isolated from either FF tissue (normal) or from FFPE blocks was subjected to qRT-PCR using iScript Select cDNA Synthesis Kit (Bio-Rad) and SsoFast EvaGreen Supermix (Bio-Rad) according to manufacturers' instructions. Two tRNAs showing prognostic significance and a FC of > 2.0 were chosen for validation using total RNA from nine normal samples and 44 tumor samples. These samples were also used for sequencing experiment. Random primers were used for reverse transcription. Primers for analyzing chr6.tRNA50-SerAGA and chr6.tRNA51-SerTGA were designed with Primer3 software. The sequence of the primer pairs are as follows:

chr6.tRNA50-SerAGA -F: 5'-TAGTCGTGGCCGAGTGGTTA-3',

chr6.tRNA50-SerAGA -R: 5'-GGAAACCCCAATGGATTTCTA-3'; and for

chr6.tRNA51-SerTGA -F: 5'-TAGTCGTGGCCGAGTGGTTA-3',

chr6.tRNA51-SerTGA -R: 5'-GAAACCCCAATGGATTTCAA-3'. GAPDH served as the loading control. Primers for analyzing GAPDH by qRT-PCR are described elsewhere [31]. All experiments for qRT-PCR were done in triplicates, data was analyzed using the $2^{-\Delta\Delta Ct}$ method [32], and results are shown as fold induction of tRNAs.

6.2.2 Genomic distribution of tRNAs, identification of regulatory RNAs embedded within tRNAs and their roles in gene regulation

With the objective of identifying the possible sites of origin of tRNAs, I overlapped the genomic co-ordinates of all the tRNAs profiled (n = 571) with the genomic co-ordinates of mRNAs and lncRNAs using PGS.

Previous studies have reported that tRNAs may also act as reservoirs for other regulatory RNAs such as miRNAs [13] and piRNAs [17]. Therefore the genomic co-ordinates of all 571 tRNAs were overlapped with the genomic co-ordinates of mature miRNAs and piRNAs. Since miRNAs and piRNAs are considered as master regulators of gene expression, potential mRNA targets were identified from gene (mRNA) expression dataset that was available in house (GEO accession ID: GSE22820) [33]. The dataset included 10 normal breast tissues (obtained from reduction mammoplasty) and 141 breast tumor tissues. PGS v 6.6 was used for all the analysis. Raw data was quantile normalized and log2 transformed, and mRNAs exhibiting FC > 2.0 and FDR $\leq$ 0.05 were identified as DE using ANOVA.

mRNA targets for piRNAs embedded within tRNAs were identified using miRanda v 3.3a. The piRNAs identified to be within the tRNAs were found to be up-regulated in tumor tissues, relative to normal breast tissues [19]. Therefore, fasta sequences of the 3'UTRs of all the down-regulated genes downloaded from Ensembl database (GRCh37) [34] and fasta sequences for piRNAs (whicch were all up-regulated in the study) were accessed from the piRNA bank (hg 19) [35]. mRNA-piRNA pairs showing sequence complementarity, with alignment score $\geq$ 170 and energy score $\leq$ -20 kcal/mol were identified. The targets thus identified were interrogated for gene ontology classifications to gain functional insights. Gene ontology classification was performed using PGS and gene ontology terms (biological process) showing enrichment score $\geq$ 1.3 and a p-value $\leq$ 0.05 were considered.

6.3 Results

All the profiling results of tRNAs and the details on identifying differentially expressed tRNAs are summarized in chapter 3 (3.5).

Overall, 76 tRNAs were DE with FC > 2 and FDR cut off 0.05 (Appendix Table 9.1) and all 76 tRNAs were up-regulated in tumor tissue compared to normal tissue, indicating a global up-regulation of tRNAs in BC.

6.3.1 tRNAs are associated with breast cancer prognosis

Two approaches (CC and CO) were adopted to identify tRNAs as potential prognostic markers for BC (*vide* methods, 2.5.1).

Case-control approach:

In the CC approach, survival analysis was restricted to 76 DE tRNAs that were subjected to univariate Cox proportional hazards regression model followed by permutation test. I found three tRNAs (chr6.tRNA5-SerAGA, chr6.tRNA50-SerAGA and chr6.tRNA51-SerTGA) to be associated with OS, with a permutation p-value ≤ 0.1 (Table 6.1). These three tRNAs were used to construct a risk score for all cases, and then the cases were dichotomized into two groups based on the ROC estimated cut-off point (1.05). Cases with a risk score ≤ 1.05 and > 1.05 were classified as low-risk and high-risk groups, respectively. Further, the risk score was adjusted for tumor stage and age at diagnosis. High-risk group patients were found to have shorter OS (hazard ratio, HR = 2.68, p=0.02, CI = 1.19 – 5.99; Table 6.2 Figure 6.1). Interestingly, none of the DE tRNAs were found to be associated with RFS.

Table 6.1 List of tRNAs significant for Overall survival

tRNA ID	Univariate Cox p-value	Permuted p-value
Chr6.tRNA166-AlaAGC	0.02	0.04
Chr17.tRNA10-GlyTCC	0.04	0.05
Chr6.tRNA147-SerAGA	0.04	0.06
Chr6.tRNA145-SerAGA	0.04	0.06
Chr6.tRNA5-SerAGA	**0.06**	**0.07**
Chr16.tRNA2-ArgCCT	0.04	0.08
Chr6.tRNA50-SerAGA	**0.07**	**0.09**
Chr12.tRNA8-AlaTGC	0.08	0.09
Chr6.tRNA148-SerTGA	0.07	0.09
Chr6.tRNA172-SerTGA	0.07	0.09
Chr6.tRNA143-LysTTT	0.06	0.09
Chr14.tRNA2-LeuTAG	0.07	0.09
Chr6.tRNA51-SerTGA	**0.08**	**0.09**
Chr9.tRNA4-ArgTCT	0.06	0.10

Table 6.1: Two approaches were adopted to select the set of tRNAs for survival analysis. In the CC approach and CO approach, 76 DE tRNAs and 216 tRNAs (retained after filtering for read counts) were selected for Univariate Cox proportional hazards regression model (outcome: OS), followed by permutation test. Table 6.1 includes OS significant tRNAs (permuted p-value ≤ 0.1) from both the approaches (n = 3 in CC and n = 14 in CO). The CO approach also included the tRNAs that were significant in the CC approach, which are indicated in red color.

Figure 6.1 tRNA Kaplan-Meier plot for Overall Survival (Case-control)

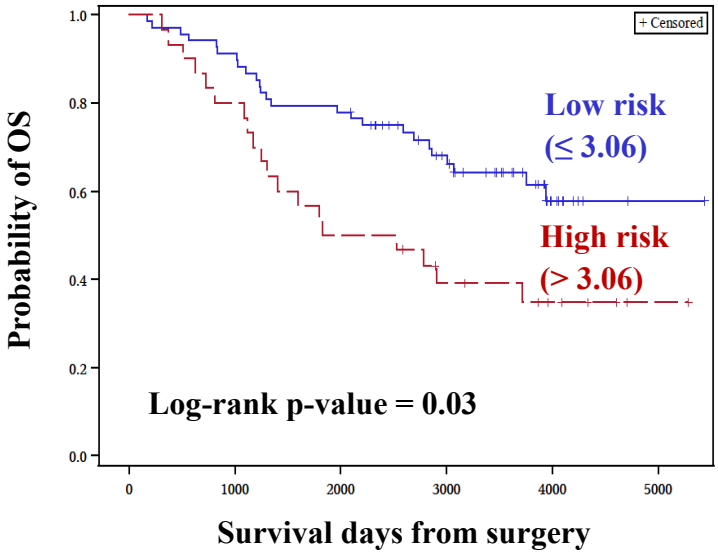


Figure 6.1: Probability of OS is plotted over time and the Kaplan-Meier plot indicate that in CC, relative to low-risk group, patients belonging to high-risk group are associated with poorer OS.

Table 6.2 Univariate and Multivariate results of tRNAs for Case-control approach

Parameter	Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.39 (1.07-5.33)	0.03	2.68 (1.19-5.99)	0.02
Tumor stage	0.40 (0.21-0.78)	0.01	0.50 (0.25-1.01)	0.05
Tumor grade	2.01 (1.04-3.89)	0.04		
Age at diagnosis	1.06 (1.02-1.09)	0.001	1.05 (1.02-1.09)	0.002
TNBC status	0.99 (0.50-1.95)	0.98		

HR = Hazard Ratio; CI = Confidence Interval; TNBC = Triple Negative Breast Cancer

Table 6.2: Risk scores were constructed from the three tRNAs (significant for OS) identified from CC. Patients were dichotomized into low and high risk groups based on ROC estimated cut-off point. Univariate Cox analysis was run for risk score and other clinical variables (included in the table). Risk score was further adjusted for potential confounders and was found to be significant ($p < 0.05$). Patients belonging to high-risk group were associated with poorer OS ($HR > 1$).

Case-only approach:

571 tRNAs were profiled from tumor tissues alone, of which, 216 were retained with ≥ 10 read counts in at least 90% of tumor samples. The dataset was RPKM normalized and adjusted for batch effects. From the 216 tRNAs (treated as continuous variables), 14 tRNAs each were significant for OS (Table 6.1) and RFS (Table 6.3), respectively in the permutation test, following Univariate Cox analysis. The 14 tRNAs significant for OS, included the three tRNAs that were significant in the CC approach. The estimated optimal cut-off point for defining risk groups was 7.23, and patients were stratified into low-risk (≤ 7.23) and high-risk groups (> 7.23) for OS. Similar to the CC approach, high-risk group was found to be associated with shorter OS (HR = 2.78, p = 0.0008, CI = 1.53 – 5.07, Table 6.4, Figure 6.2). In contrast to the CC approach, 14 tRNAs were found to be significant for RFS (Table 6.3). A risk score cut-off point of -3.11 separated cases into two survival groups and the high-risk group was found to be associated with shorter RFS (HR = 1.86, p = 0.02, CI = 1.10 – 3.13, Table 6.4, Figure 6.3). For both OS and RFS, risk score was found to be significant after adjusting for confounders (tumor stage, grade and age at diagnosis for OS and tumor stage for RFS).

Table 6.3 List of tRNAs significant for Recurrence Free Survival

tRNA ID	Univariate Cox p-value	Permuted p-value
Chr6.tRNA166-AlaAGC	0.03	0.03
Chr1.tRNA80-GluCTC	0.05	0.04
Chr1.tRNA77-GluCTC	0.05	0.04
Chr6.tRNA87-GluCTC	0.07	0.06
Chr1.tRNA74-GluCTC	0.07	0.06
Chr1.tRNA71-GluCTC	0.07	0.06
Chr1.tRNA59-GluCTC	0.08	0.06
Chr6.tRNA77-GluCTC	0.08	0.07
Chr1.tRNA118-HisGTG	0.1	0.08
Chr6.tRNA152-ValCAC	0.13	0.08
Chr1.tRNA116-GluCTC	0.11	0.09
Chr2.tRNA19-GlyGCC	0.12	0.09
Chr6.tRNA128-GlyGCC	0.11	0.09
Chr1.tRNA133-GlyCCC	0.12	0.09

Table 6.3: Two approaches were adopted to select the set of tRNAs for survival analysis. In the CC approach and CO approach, 76 DE tRNAs and 216 tRNAs (retained after filtering for read counts) were selected for Univariate Cox proportional hazards regression model (outcome: RFS), followed by permutation test. Table 6.3 includes RFS significant tRNAs (permuted p-value ≤ 0.1) from CO approach (n = 14). None of the tRNAs were identified as associated with RFS from the CC approach.

Figure 6.2 tRNA Kaplan-Meier plot for Overall Survival (case-only)

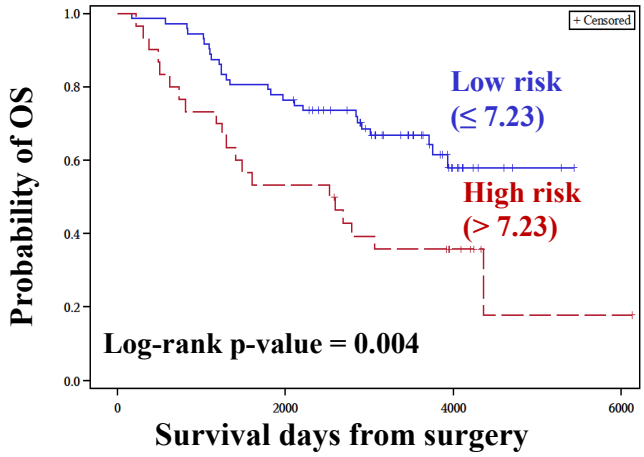
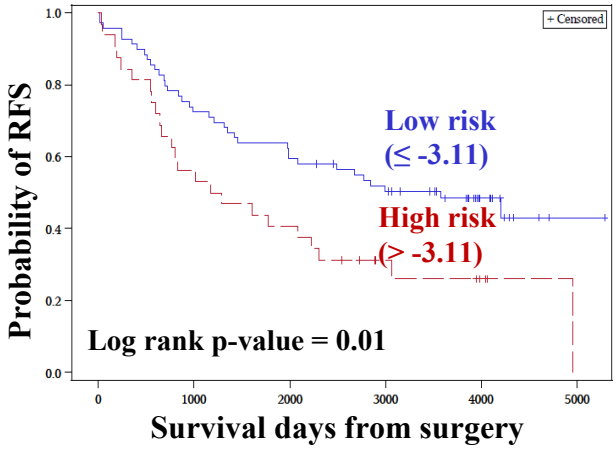


Figure 6.3 tRNA Kaplan-Meier plot for Recurrence free Survival (case-only)



Figures 6.2 and 6.3: Probability of OS (Figure 6.2) and RFS (Figure 6.3) is plotted over time and the Kaplan-Meier plots indicate that relative to low-risk group, patients belonging to high-risk group are associated with poorer OS and RFS.

Table 6.4 Univariate and Multivariate results of tRNAs for Case-only approach

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.33 (1.29-4.18)	0.01	2.78 (1.53-5.07)	0.001	1.89 (1.13-3.19)	0.02	1.86 (1.10-3.13)	0.02
Tumor stage	0.40 (0.21-0.78)	0.01	0.46 (0.23-0.93	0.03	0.38 (0.20-0.71)	0.003	0.39 (0.21-0.73)	0.003
Tumor grade	2.01 (1.04-3.89)	0.04	2.49 (1.26-4.93)	0.01	1.58 (0.92-2.74)	0.10		
Age at diagnosis	1.06 (1.02-1.09)	0.001	1.05 (1.02-1.09)	0.001	1.02 (0.99-1.05)	0.21		
TNBC status	0.99 (0.50-1.95)	0.98			0.84 (0.45-1.55)	0.58		

HR = Hazard Ratio; CI = Confidence Interval; TNBC = Triple Negative Breast Cancer

Table 6.4: Risk scores were constructed from the 14 tRNAs (significant for OS and RFS) identified from CO. Patients were dichotomized into low and high risk groups based on ROC estimated cut-off point. Univariate Cox analysis was run for risk score and other clinical variables (included in the table). Risk score was further adjusted for potential confounders and was found to be significant ($p < 0.05$). Patients belonging to high-risk group were associated with poorer OS (left panel) and RFS (right panel) with HR > 1.

6.3.2 tRNAs prognostic of overall survival are validated in an external dataset

The batch adjusted normalized counts for tRNAs associated with OS (identified in the CO approach) were extracted from TCGA dataset. Similar to the discovery cohort, risk scores were constructed for every sample and the samples were dichotomized into low and high-risk groups based on the cut-off point (-0.9) estimated using ROC. In the multivariate setting, the risk score was adjusted for tumor stage and age at diagnosis. Statistical significance obtained ($p = 0.15$) for the risk score indicated a trend similar to the original study (similar direction and magnitude of effect) but did not meet imposed nominal p-value threshold of $p < 0.05$. Overall, the results from external cohort were

supportive of the original study findings that tRNAs are potential prognostic factors; high-risk group was associated with poorer OS (HR = 1.97, p = 0.15, CI = 0.79 – 4.95, Table 6.5, Figure 6.4).

Figure 6.4 Kaplan-Meier plot for Overall Survival (External/TCGA dataset)

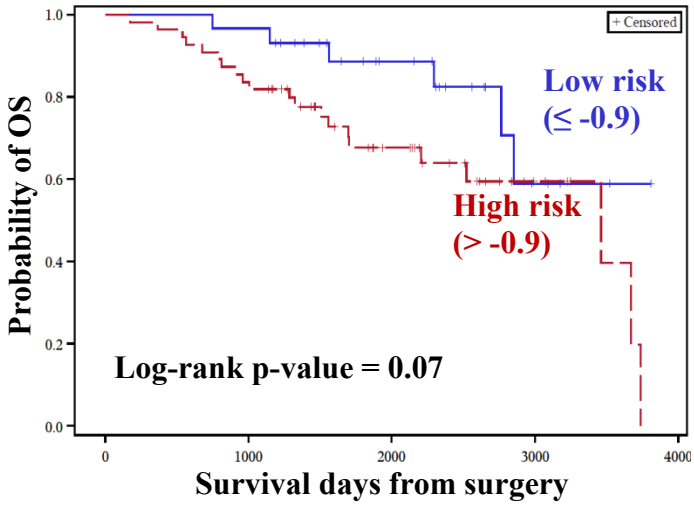


Figure 6.4: Probability of OS is plotted over time for the TCGA dataset. Kaplan-Meier plot indicates that, relative to low-risk group, patients belonging to high-risk group are associated with poorer OS, similar to discovery cohort.

Table 6.5 Univariate and multivariate analysis of tRNAs for Overall Survival (External/TCGA dataset)

Parameter	Univariate analysis		Multivariate analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Risk score	2.28 (0.92 – 5.66)	0.08	1.97 (0.79 – 4.95)	0.15
Tumor stage	0.32 (0.13 – 0.78)	0.01	0.29 (0.11 – 0.74)	0.009
Age at diagnosis	1.03 (1.003 – 1.06)	0.03	1.03 (1.01 – 1.06)	0.02
TNBC status	0.63 (0.19 – 2.12)	0.46		

HR = Hazard ratio; CI = Confidence interval; TNBC = Triple Negative Breast Cancer

Table 6.5: Risk score was constructed using the 14 tRNAs (significant for OS) for all the 84 samples accessed from the TCGA dataset. Patients were dichotomized into low and high risk groups based on ROC estimated cut-off point. Univariate Cox analysis was run for risk score and other clinical variables (included in the table). Risk score was further adjusted for potential confounders and was found to be significant with p = 0.15). Similar to the discovery cohort, patients belonging to high-risk group were associated with poorer OS (HR > 1).

6.3.3 Relative expressions of chr6.tRNA50-SerAGA and chr6.tRNA51-SerTGA are validated using qRT-PCR

Two representative tRNAs, chr6.tRNA50-SerAGA and chr6.tRNA51-SerTGA, exhibiting a fold change of 2.56 and 2.61, respectively in NGS platform, were validated using qRT-PCR. Both tRNAs were found to be up-regulated in tumors relative to normal tissues in qRT-PCR experiments (Figure 6.5).

Figure 6.5 qRT-PCR validation of up-regulated tRNAs

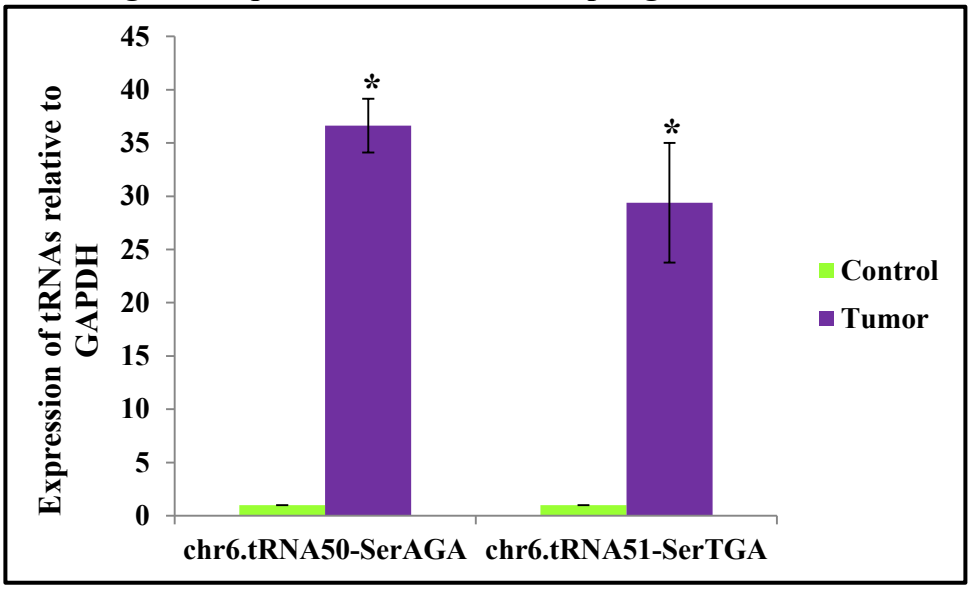


Figure 6.5: The expressions of two prognostically significant tRNAs that were also differentially expressed were validated using qRT-PCR with GAPDH as the internal normalizer. Both the tRNAs are up-regulated in breast tumor, relative to normal (control) tissues, conforming the findings from NGS experiment. * = $p < 0.05$.

6.3.4 tRNAs harbor regulatory RNAs and thus contribute to gene regulation

Genomic origins (distinct genes or intergenic or intragenic regions) of tRNAs are not well understood . However, in this study, I observe that a fraction of tRNAs appears to originate from the intronic regions of protein coding or non-protein coding genes. For instance, when I mapped the genomic co-ordinates of the 571 profiled tRNAs to the genomic co-ordinates of messenger RNAs (mRNAs) and long non-coding RNAs (lncRNAs), I observed that ~ 15% (n = 86) of the tRNAs were embedded within the intronic regions of mRNAs and ~ 12% (n = 66) were embedded within the introns of lncRNAs (Appendix Table 9.4).

Since we now understand that tRNAs may also act as a reservoir for microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs), the genomic co-ordinates of the 571

tRNAs were mapped to the genomic co-ordinates of mature miRNAs and piRNAs. 45 tRNAs were observed to harbor piRNAs (Appendix Table 9.5) and one tRNA was found to harbor a miRNA (Appendix Table 9.5). The identified piRNAs were subsequently interrogated for differential expression using data generated from our previous studies (Chapter 5) [19]. Nine piRNAs (from among the 45 piRNAs annotated to tRNAs) were found to be up-regulated (Table 6.6). The lone miRNA observed to be within the genomic co-ordinates of a tRNA was not found to be differentially expressed.

Further, to understand the contribution of piRNAs (thereby the tRNAs) to gene regulation, I first identified mRNA targets based on (i) the complementary sequences shared by piRNAs and the 3'UTR of mRNAs and (ii) the reciprocal expression patterns between piRNAs and mRNAs. Since the nine piRNAs were found to be up-regulated, 2241 genes which were found to be down-regulated in tumor tissues (gene expression dataset), were considered as potential targets for the nine piRNAs. However, when filtered for stringent alignment and energy scores (Table 6.6), 76 targets (genes) were identified. To understand the functional relevance of the identified targets, gene ontology classification was performed and the identified gene ontology terms (biological processes) are summarized in Table 6.7. The identified targets were found to be involved in key tumorigenic pathways, including apoptosis and angiogenesis.

Table 6.6 List of gene targets identified by piRNAs embedded within tRNAs

tRNA ID (Fold change)	piRNA ID (Fold change)	mRNA targets (Down-regulated)
chr1.trna68-GlyGCC (1.88)	hsa_piR_000291 (1.71)	TNKS, ZC3H6, ZHX3
chr2.trna19-GlyGCC (1.99)	hsa_piR_000765 (1.85)	SCN2B, SH3TC2, SEMA6D, SLC16A4, SYNPO, TMTC1, TSHZ2, TIFA, TRPM3, WFIKNN2, ZSCAN12, UBQLNL, APCDD1, CNR1, CES2
chr6.trna13-LysCTT (14.79)	hsa_piR_000794 (1.94)	RRAD, SLC2A4, SEMA3E, RPL18, ZNF366, WSCD1, B3GAT1, CACNA1B, CES2
chr6.trna5-SerAGA (2.46)	hsa_piR_015249 (2.42)	NONE
chr6.trna87-GluCTC (1.35)	hsa_piR_017716 (1.51)	SEMA3G, SCARA3, SIRPA, RSPO1, SLC23A2, RPS9, SLC34A2, ST8SIA2, TMCC3, TLN2, TNFSF12, TRIM2, TIFA, ZNF395, TXNRD2, VPRBP, ADAM11, ACVR1C, ANGPTL4, ACACB, ALS2CL, APOL4, ALPL, ARID5A, ATP13A4, ACSM1, CLEC4M, CLIP3, CCDC120, CCDC38
chr19.trna8-SeC(e)TCA (18.15)	hsa_piR_019912 (16.64)	SDK2, SYNPO
chr12.trna13-AlaTGC (1.14)	hsa_piR_020485 (1.11)	SLC2A4, SEC63, TMEM87A, USP31, VPS13A, AKR1C1, ABCG5, ALG9
chr2.trna3-AlaAGC (1.87)	hsa_piR_020496 (1.87)	ALG9
chr5.trna15-ValAAC (9.57)	hsa_piR_020829 (9.58)	SCN2B, SACS, RYR1, SNCAIP, WNT5B, ARHGAP26, CAPN6, CD34

Table 6.6: 45 piRNAs were found to be embedded within tRNAs, of which nine piRNAs were found to be differentially expressed. Since these 9 piRNAs were up-regulated, potential targets were identified from the genes that were down-regulated in breast tumor tissues. A total of 76 gene targets were identified for the 9 piRNAs.

Table 6.7 Gene ontology classification for the piRNA targets

Gene ontology classification	mRNA targets	piRNAs regulating mRNA target expression
Regulation of angiogenesis	SEMA3E, TNFSF12, ANGPTL4, CD34	hsa_piR_000794, hsa_piR_017716, hsa_piR_020829
Apoptotic nuclear changes	ACVR1C	hsa_piR_017716
Fat cell differentiation	SLC2A4, CLIP3, WNT5B	hsa_piR-000794, hsa_piR_020485, hsa_piR_017716, hsa_piR_020829
Regulation of Wnt signaling pathway	TNKS, APCDD1, RSPO1, WNT5B	hsa_piR_000291, hsa_piR_000765, hsa_piR_017716, hsa_piR_020829
Doxorubicin and Daunorubicin metabolic process	AKR1C1	hsa_piR_020485
Negative regulation of intracellular estrogen receptor signaling pathway	ZNF366	hsa_piR_000794
Progesterone metabolic process	AKR1C1	hsa_piR_020485
Hematopoietic stem cell proliferation	CD34	hsa_piR_020829

Table 6.7: Representative gene ontology terms with enrichment score > 1.3 and p-value < 0.05 are listed. Each row in columns two and three represent the mRNA targets involved in the functions and the corresponding piRNAs predicted to bind to these targets.

6.4 Discussion

This is the first study to profile tRNAs on a genome wide scale using NGS and to identify their prognostic significance for BC. 571 tRNAs were profiled and I found that 14 tRNAs each, were associated with OS and RFS. Amongst these, one tRNA was found to be associated with both OS and RFS. The results also showed similar direction of effect in an external dataset, thereby strengthening the study findings.

This study provides proof of principle experiments in support of the idea that a comprehensive tumor profiling of tRNAs will offer much-needed insights in to new biomarkers for BC prognosis. The two approaches used in the study, CC and CO (do not depend on controls used), are widely accepted means to identify markers of prognostic significance. Although, it is not common to adopt both approaches in a single study, both the approaches have been attempted in this study to compare and understand their similarities and differences, in terms of number of signatures and/or the unique or common signatures captured. As anticipated, the number of prognostically significant tRNAs identified were higher in a CO approach since the number of tRNAs interrogated for survival analysis was also higher. In the case of OS, three tRNAs were found to be significant in CC approach, while 14 tRNAs were identified in CO approach. No tRNAs were associated with RFS in CC approach whereas 14 tRNAs were found to be significant from a CO approach. Therefore, adopting a CO approach not only offers a larger dataset to probe for markers but is also a better option to understand the importance of molecules which would have otherwise been missed in a CC approach that focusses only on DE tRNAs.

A stringent filtering criterion was adopted, that enabled to identify tRNAs present in high amounts, and in most, if not, all of the samples (highly expressed and most frequently expressed). This is one way to improve the chance of reproducibility of the obtained signatures. Indeed, all the 14 tRNAs significant for OS had read counts ≥ 10 in at least 90% of the samples (except one, which had ≥ 10 read counts in 87% of the samples) in the external/TCGA dataset, and therefore the overall expression levels were considered as comparable to the discovery dataset. Results of survival analysis from

TCGA dataset showed a similar direction of effect; patients belonging to high-risk group were associated with poorer OS, validating the findings from the discovery cohort. The risk score, however, did not reach statistical significance due to the limited sample size and number of events (death) in the cohort, a finding consistent in independent biomarker studies when TCGA dataset was considered for the studies outlined in this thesis [30]. Recurrence events reported for the TCGA dataset are lower than OS and hence the data was not amenable for RFS analysis.

To build a model for multivariate analysis, I did not include individual tRNA molecules identified from the univariate analysis but constructed a composite risk score using these RNAs for the following reasons: (i) a complex interplay of biomolecules exists, where each molecule contributes significantly towards a phenotype; (ii) several of the tRNAs identified are highly correlated (Appendix Tables 9.6 and 9.7) and the pattern of correlation is more pronounced for tRNA isoacceptors ($r = > 0.9$). While this was expected for isoacceptors, it was also interesting to observe fairly high correlation ($r = > 0.8$) between tRNA genes coding for Ser and Leu (specific reason not known). This problem of collinearity which generally leads to spurious associations [36] of the variables with the outcomes was also overcome by constructing a risk score, which is usually not affected by correlated variables.

Recent studies have highlighted the importance of tRNAs as a source for other regulatory RNAs such as piRNAs [17] and miRNAs [13], which act as master regulators of gene expression. To this end, I observed that 46 tRNAs potentially harbor these regulatory RNAs. I also identified that among these 46 regulatory RNAs (piRNAs, miRNAs), nine piRNAs were DE. The nine piRNAs were predicted to target a total of 76

mRNA targets from gene expression dataset obtained from breast tissues. Since these targets are obtained from breast tissues, this dataset may serve as proxy for functional validation. Gene ontology classifications of these targets were enriched for key tumorigenic pathways such as angiogenesis, apoptosis and stem cell maintenance (Table 6.7).

Although I have identified a potential indirect role of tRNAs in breast tumorigenesis, yet one needs to confirm if these tRNAs are indeed giving rise to these regulatory RNAs or if a portion of tRNA merely share sequence similarity to piRNAs. One level of evidence from this study to say that these piRNAs may be embedded within the tRNAs is that the expression of piRNAs and tRNAs were found to be in the same direction. All the nine piRNAs and their corresponding host tRNAs were up-regulated in breast tumor tissues (Table 6.6). A series of experiments are needed to confirm the piRNA origins to tRNAs: These include (i) the expression studies to correlate piRNA and host genes showing similar direction of expression, (ii) demonstrate interactions of piRNAs with PIWI proteins, which are the drivers of piRNA biogenesis, (iii) demonstrate direct interaction between the piRNAs and the identified mRNA targets thrugh luciferase expression systems, and (iv) assess potential functions in cellular activities (apoptosis, cell migration, cell proliferation etc) using cell based assays.

Frequently used methods to estimate the cut-off point for patient stratification into two survival groups are – median cut-off point of the risk score and ROC based cut-off point. While calculating the median is the most commonly adopted method, this cut-off point is arbitrary [37] and does not take into account the sensitivity and specificity of the estimated cut-off point. Conversely, ROC based estimation considers these and is a more

reliable measure for cut-off point estimation [38]. ROC based estimation was therefore used to determine the cut-off point for patient stratification. Overall, this study has satisfied the parameters set by REMARK guidelines [39] for biomarker discovery and validation.

6.5 Conclusions

This is the first study to comprehensively profile tRNAs using NGS and to understand their contribution to BC prognosis. Despite the technical challenges involved in sequencing tRNAs, this study has demonstrated a near complete capture of all the annotated tRNAs using the data from the adopted NGS platform. Results from this study also indicate that tRNAs may emerge as promising prognostic biomarkers for BC and an observation of the same trends of association with BC prognosis in an external dataset, reaffirms the initial study findings. However, it remains to be seen if these tRNA molecules may perform better as stand-alone biomarkers or if these can complement the existing prognostic markers for BC. Confirmation of the processing of tRNAs to other regulatory RNAs may add a new dimension to the existing knowledge on tRNAs, which may also be beneficial for therapeutic purposes. I believe that the findings from the current study will encourage more researchers to contribute to delineate the fine molecular mechanisms. Although much remains to be ascertained regarding the various aspects of tRNAs, deeper exploration into this class of RNAs may help us better appreciate the hitherto unexplored biological consequences of these RNAs.

6.6 References

1. Palazzo AF, Gregory TR. The case for junk DNA. PLoS Genetics 2014;10:e1004351.

2. Cao JN. The functional role of long non-coding RNAs and epigenetics. Biol Proc Online 2014;16:11.

3. Ronnau C, Verhaegh GW, Luna-Velez M, Schalken JA. Noncoding RNAs as novel biomarkers in prostate cancer. Biomed research international 2014;2014:591703.

4. Berindan-Neagoe I, Monroig PD, Pasculli B, Calin GA. MicroRNAome genome: A treasure for cancer diagnosis and therapy. CA-A Cancer journal for clinicians 2014;64:311-36.

5. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955-64.

6. Wek SA, Zhu S, Wek RC. The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. Mol Cell Biol 1995;15:4497-4506.

7. Mei YD, Yong J, Liu HT, Shi YG, Meinkoth J, Dreyfuss G, et al. tRNA binds to cytochrome c and inhibits caspase activation. Mol Cell 2010;37:668-78.

8. Ciechanover A, Wolin SL, Steitz JA, Lodish HF. Transfer RNA is an essential component of the ubiquitin- and ATP-dependent proteolytic system. Proc Natl Acad Sci U S A 1985:1341.

9. Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 2009;23:2639-49.

10. Wang QR, Lee I, Ren JP, Ajay SS, Lee YS, Bao XY. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. Molecular therapy 2013;21:368-79.

11. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep sequencing of serum small RNAs identifies patterns of 5' tRNA half and YRNA fragment expression associated with breast cancer. Biomark Cancer 2014;6:37-47.

12. Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. Cell 2015;161:790-802.

13. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 2013;110:1404-9.

14. Olvedy M, Scaravilli M, Hoogstrate Y, Visakorpi T, Jenster G, Martens-Uzunova E. A comprehensive repertoire of tRNA-derived fragments in prostate cancer. Oncotarget 2016.

15. Kovalchuk O, Filkowski J, Meservy J, Ilnytskyy Y, Tryndyak VP, Chekhun VF, et al. Involvement of microRNA-451 in resistance of the MCF-7 breast cancer cells to chemotherapeutic drug doxorubicin. Mol Cancer Ther 2008;7:2152-2159.

16. Wang B, Komers R, Carew R, Winbanks CE, Xu B, Herman-Edelstein M, et al. Suppression of microRNA-29 expression by TGF-ß1 promotes collagen expression and renal fibrosis. J Am Soc Nephrol 2012;23:252-265.

17. Keam SP, Young PE, McCorkindale AL, Dang THY, Clancy JL, Humphreys DT, et al. The human piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. Nucleic Acids Res 2014;42:8984-8995.

18. Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-9910.

19. Krishnan P, Ghosh S, Graham K, Mackey J, Kovalchuk O, Damaraju S. Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer. Oncotarget 2016.

20. Bjornsti MA, Houghton PJ. Lost in translation: Dysregulation of cap-dependent translation and cancer. Cancer cell 2004;5:519-523.

21. Chen W, Böcker W, Brosius J, Tiedge H. Expression of neural BC200 RNA in human tumours. J Pathol 1997;183:345-351.

22. Marshall L, White RJ. Non-coding RNA production by RNA polymerase III is implicated in cancer. Nature Reviews Cancer 2008:911.

23. Winter AG, Sourvinos G, Allison SJ, Tosh K, Scott PH, Spandidos DA, et al. RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. Proc Natl Acad Sci U S A 2000:12619.

24. Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. RNA 2013;19:461-466.

25. Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC. A soluble ribonucleic acid intermediate in protein synthesis. J Biol Chem 1958;231:241-257.

26. Hoagland MB, Keller EB, Zamecnik PC. Enzymatic carboxyl activation of amino acids. J Biol Chem 1956;218:345-358.

27. Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res 2009;37:7268-7280.

28. Bertucci F, Birnbaum D. Reasons for breast cancer heterogeneity. J Biol 2008;7:6-6.

29. Ribelles N, Perez-Villa L, Jerez JM, Pajares B, Vicioso L, Jimenez B, et al. Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. Breast Cancer Res 2013;15:R98-R98.

30. Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics 2015;16:735-735.

31. Wang B, Khachigian LM, Esau L, Birrer MJ, Zhao X, Parker MI, et al. A key role for early growth response-1 and nuclear factor-kappaB in mediating and maintaining GRO/CXCR2 proliferative signaling in esophageal cancer. Mol Cancer Res 2009;7:755-764.

32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 2001;25:402-408.

33. Germain DR, Graham K, Glubrecht DD, Hugh JC, Mackey JR, Godbout R. DEAD box 1: A novel and independent prognostic marker for early recurrence in breast cancer. Breast Cancer Res Treat 2011;127:53-63.

34. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl genomes 2016: More genomes, more complexity. Nucleic Acids Res 2016;44:D574-D580.

35. Sai Lakshmi S, Agrawal S. piRNABank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173-D177.

36. Tu YK, Kellett M, Clerehugh V, Gilthorpe MS. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. Br Dent J 2005;199:457-61.

37. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: International survey. BMJ 2000;320:1240-1243.

38. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013;4:627-635.

39. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Guidelines: REporting recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer 2005;41:1690-1696.

7 Profiling of Small Nucleolar RNAs by Next Generation Sequencing: Potential New Players for Breast Cancer Prognosis

7.1 Introduction

Breast cancer (BC) is a complex polygenic phenotype [1] characterized by molecular and histological heterogeneity [2]. Although the diagnostic and prognostic factors related to breast cancer outcomes have become increasingly refined, there remains a need to improve the quality of life for BC patients. Optimal management of BC is challenging due to the varied treatment response patterns exhibited by patients undergoing similar treatment regimens [3,4]. However, the available treatment modalities might be better applied if we could stratify treatment responders from non-responders, which may eventually help in improving survival and quality of life. Although estrogen, progesterone and human epidermal growth factor receptors are routinely used as prognostic markers, in addition to tumor and patient related factors, these indices remain as imperfect estimators for risk of recurrence and/or death [5]. Therefore, there is an ongoing search in BC for better prognostic markers. .

With the discovery of new classes of small non-coding RNAs, their functions are ever expanding. Among the many small non-coding RNAs identified so far, microRNAs (miRNAs) are well established as global regulators of gene expression [6-10], that have also been studied comprehensively as biomarkers for various cancer types [11-16]. On the contrary, one of the lesser studied classes of small non-coding RNAs is the group of small nucleolar RNAs (snoRNAs), which are approximately 60–300 nt in length [17]. snoRNAs often originate within the nucleolus of a cell and are mostly encoded within the

A version of this chapter has been resubmitted to PLoS One after revisions.
I thank Dr. Carol Cass for hercritical reading of the manuscript.

intronic regions of protein-coding or non-protein coding genes such as long non-coding RNAs, or are independently transcribed from the intergenic regions [18]. Predominantly, they are classified into two groups: SNORAs, containing H/ACA box; and SNORDs, containing C/D box [19]. scaRNAs or small Cajal body RNAs, can also be classified as snoRNAs [20]. snoRNAs are often involved in ribosomal RNA (rRNA) maturation and biogenesis and is also involved in modifications of other RNAs such as rRNAs, transfer RNAs (tRNAs) and small nuclear RNAs (snRNAs). Specifically, SNORAs are involved in pseudouridylation through their association with dyskerin protein and SNORDs, along with fibrillarin proteins, are involved in methylation. Nevertheless, not all snoRNAs have defined functions and these are called "orphan snoRNAs" [20].

While the snoRNAs are largely recognized for playing housekeeping roles, emerging evidence suggests that dysregulation of snoRNAs occurs in various diseases. The first indication on the pathological importance of snoRNAs arose from the observation that a genetic locus containing snoRNAs was deleted in Prader Willi syndrome, a neurodevelopmental genetic disorder [21]. snoRNA deregulation has been observed in metabolic stress disorder [22] and in chronic conditions such as chronic lymphocytic leukemia [23], hepatocellular carcinoma [24], colorectal cancer [25] and endometrial cancer [26], among others. Extending these observations, their roles as diagnostic and prognostic biomarkers have also been studied for several cancer types including colorectal cancer [25] and lung cancer [27,28]. Although reports by Dong et al [29] and Su et al [30] have implicated the importance of snoRNAs in breast carcinogenesis, a comprehensive understanding of snoRNAs as prognostic markers for BC is still lacking. snoRNAs are also beginning to be understood as indirect regulators of gene expression.

snoRNAs may get processed to other smaller regulatory RNAs such as miRNAs and piwi-interacting RNAs (piRNAs), which are well known as post-transcriptional gene regulators [17,31,32].

We hypothesized that deregulation of snoRNAs contributes to inter-individual differences in BC trajectory and eventual outcomes. In this study, we investigated the potential of snoRNAs as prognostic markers for BC, focusing on overall survival (OS) and recurrence free survival (RFS). We have also explored the possible regulatory functions of snoRNAs. To the best of our knowledge, this is the first study to identify snoRNAs as potential independent prognostic markers for BC.

7.2 Materials and methods

I have provided the clinical characteristics of the discovery cohort samples used for the study and the methods involved in total RNA isolation, sequencing and data analysis, including the methods for identifying prognostic markers in chapter 2 (sections 2.2.1, 2.3, 2.4 and 2.5.1).

All methods pertaining to snoRNAs exclusively are elaborated below.

7.2.1 Validation of snoRNA expression using qRT-PCR

As is the case with other sncRNAs, qRT-PCR experiments were performed in collaboration with Dr. Kovalchuk's laboratory in University of Lethbridge and I analyzed and interpreted the data.

The expression of two representative snoRNAs showing prognostic significance (SNORD46 and SNORD89) were validated with the total RNA isolated from normal and tumor samples. Amongst the prognostically significant snoRNAs, SNORD46 and

SNORD 89 showed the highest fold changes and were therefore considered for cross platform validation. Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed using an iScript Select cDNA Synthesis Kit (Bio-Rad) and a SsoFast EvaGreen Supermix (Bio-Rad) according to manufacturers' instructions. Reverse transcription of total RNA was performed using random primers. Primers for PCR amplification of SNORD46 and SNORD89 were designed with Primer3 software and the sequences were as follows: SNORD46-F: 5'-AAT CCT TAG GCG TGG TTG TG-3', SNORD46-R: 5'-ATG ACA AGT CCT TGC ATT GG-3'; and SNORD89-F: 5'-GAC AAG AAA AGG CCG AAT TG-3', SNORD89-R: 5'-CAT GGA GAG CAA ACT GCT GA-3'. RNU6-2 served as loading control and the primer sequences were RNU6-2-F: 5'-CGC TTC GGC AGC ACA TAT AC-3', RNU6-2-R: 5'-AGG GGC CAT GCT AAT CTT CT-3'. All assays were done in triplicates, data was analyzed using the $2^{-\Delta\Delta Ct}$ method [33], and results are shown as fold induction of snoRNAs.

7.2.2 Gene (mRNA) expression analysis

I downloaded the breast tissue gene (mRNA) expression dataset (GEO accession ID: GSE22820) which was originally generated in-house from 141 breast tumor samples and 10 normal breast tissues obtained from reduction mammoplasty using Agilent platform[13,34]. Partek Genomics Suite v6.6 served as a bioinformatic platform for gene expression analysis. The raw intensity files were quantile normalized and log2 transformed. Differentially expressed genes were identified as those exhibiting FC > 2.0 and FDR < 0.05 using one-way ANOVA.

Targets for piRNAs embedded within snoRNAs were identified using miRanda algorithm v 3.3a. I downloaded the fasta sequences of the 3' untranslated region (UTR)

of all the DE genes identified from the in-house BC gene expression dataset from Ensembl database (GRCh37, http://grch37.ensembl.org/index.html) [35] and obtained the fasta sequences of the 11 piRNAs from piRNA Bank (hg 19, http://pirnabank.ibab.ac.in/) [36]. Since piRNAs and mRNAs are known to exhibit reciprocal relationship (i.e., if a piRNA is up-regulated, the gene target is down-regulated and vice-versa), targets for the down-regulated piRNAs were interrogated from the list of up-regulated genes using miRanda. Likewise, targets for the up-regulated piRNAs were interrogated from the list of down-regulated genes. Only genes from piRNA-mRNA pairs with alignment score > 170 and energy threshold < -20 kcal/mol [37] were considered for gene ontology classification.

7.3 Results

Results of snoRNA profiling analysis and differential expression analysis are provided in chapter 3.

Explained below are the results of survival analysis and other analysis attempting to understand the biology of snoRNAs.

7.3.1 Stability of snoRNAs in FFPE tissues

Further, to investigate if snoRNAs are stable in FFPE tissues over years, we chose samples that were collected in 1996 and 2008 (the oldest and the most recently collected samples) and ran a Pearson's correlation test on the raw and normalized counts of filtered snoRNAs (n = 88). While we obtained 80.2% correlation for the raw counts, we obtained 91.3% correlation for the normalized counts, indicating the stability of the snoRNAs profiled in this study. This observation from our dataset is supported by findings from

Hall et al., who have identified snoRNAs as one of the stable molecules from FFPE tissue samples [38].

7.3.2 Thirteen snoRNAs identified with prognostic relevance for breast cancer

Case-control approach:

From tumor and normal breast tissue samples, 768 snoRNAs were identified with at least one read count. Of these, 88 snoRNAs satisfied the filtering criteria adopted (described in Chapter 2 and in other sncRNA profiling experiments in this thesis) and 40 snoRNAs exhibited dysregulation in tumor tissues (Appendix Table 9.1). The 40 snoRNAs, when considered as continuous variable and subjected to univariate Cox regression analysis, followed by permutation test, yielded five snoRNAs with prognostic significance for OS (Table 7.1) and four snoRNAs (Table 7.2) for RFS. Risk scores were constructed using these snoRNAs and cut-off points were estimated to be -3.93 and -2.75 for OS and RFS, respectively that dichotomized the samples into low-risk and high-risk groups. Risk score was considered as a dichotomous variable and results from univariate and multivariate regression analysis suggested that patients belonging to high-risk group were associated with poor outcomes and the risk score showed promise as potential independent prognostic factor for OS (HR = 3.24, CI = 1.35 – 7.77, p-value = 0.008, Figure 7.1, Table 7.3) and RFS (HR = 2.17, CI = 1.22 – 3.84, p-value = 0.008, Figure 7.2, Table 7.3).

Table 7.1 List of snoRNAs significant for Overall Survival

snoRNA ID	Univariate Cox p-value	Permuted p-value
SNORA7A	0.01	0.01
SNORD94	0.02	0.01
SNORD92	0.02	0.01
SNORD105	0.03	0.02
SNORA7B	0.03	0.03
SNORD85	0.05	0.05
SNORD46	0.05	0.05
SNORD100	0.07	0.05
SNORD104	0.07	0.05
SNORD84	0.06	0.05
SNORD82	0.06	0.05
SNORD14E	0.08	0.08

Table 7.2 List of snoRNAs significant for Recurrence Free Survival

snoRNA ID	Univariate Cox p-value	Permuted p-value
SNORD92	0.002	0.001
SNORA7B	0.02	0.01
SNORD14E	0.01	0.01
SNORD46	0.05	0.05
SNORD105	0.05	0.05
SNORD94	0.06	0.06
SNORD100	0.08	0.07
SNORD89	0.08	0.08
SNORD82	0.09	0.08
SNORA7A	0.1	0.09

Tables 7.1 and 7.2: In the CO approach, twelve and ten snoRNAs were identified for OS and RFS, respectively with permutation p-value < 0.1. The snoRNAs identified in the CO approach encompassed all the snoRNAs identified in the CC approach for both OS (n = 5) and RFS (n = 4) and are highlighted in red.

Figure 7.1 snoRNAs Kaplan-Meier plot for Overall Survival (case-control)

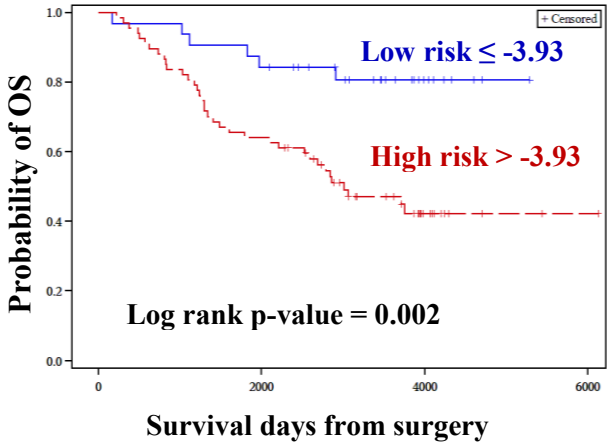
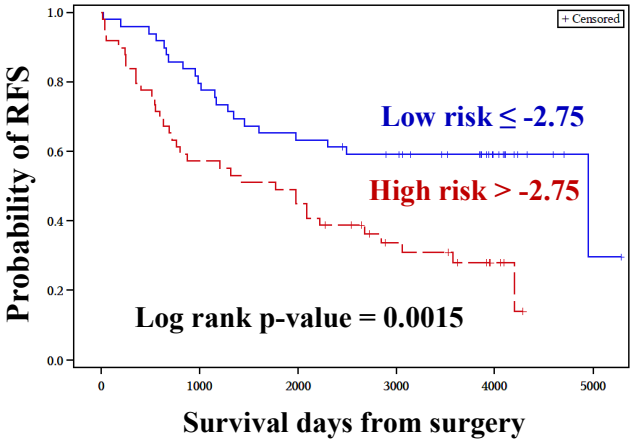


Figure 7.2 snoRNAs Kaplan-Meier plot for Recurrence Free Survival (case-control)



Figures 7.1 and 7.2: Kaplan-Meier plots for the risk scores were constructed to determine the survival differences between low–risk and high–risk groups. Significant survival differences existed between the two risk groups, as indicated by the log–rank p–values. In both OS and RFS, patients belonging to high–risk group showed poor OS (Figure 7.1) and RFS (Figure 7.2).

Table 7.3 Univariate and multivariate results of snoRNAs for case–control approach

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	P–value	HR (95% CI)	P–value	HR (95% CI)	P–value	HR (95% CI)	P–value
Risk score	3.59 (1.51–8.54)	0.004	3.24 (1.35–7.77)	0.008	2.38 (1.37–4.14)	0.002	2.17 (1.22–3.84)	0.008
Tumor stage	0.39 (0.2–0.78)	0.007			0.42 (0.22–0.78)	0.007		
Tumor grade	2.15 (1.06–4.39)	0.035	2.19 (1.07–4.52)	0.033	1.61 (0.91–2.86)	0.1		
Age at diagnosis	1.06 (1.02–1.09)	0.001	1.05 (1.02–1.09)	0.003	1.02 (0.99–1.05)	0.2		
TNBC status	0.93 (0.46–1.89)	0.83			0.76 (0.4–1.45)	0.41		

HR = Hazard ratio; CI = Confidence interval; TNBC = Triple negative breast cancer

Table 7.3: The two risk scores computed for CC approach (one for OS and one for RFS) were found to be significant in the multivariate analysis ($p < 0.05$) after adjusting for potential confounders. In both the outcomes, patients with risk scores more than the estimated optimal cut–off points were associated with poor prognosis (HR > 1).

Case-only approach:

763 snoRNAs were identified with at least one read count from all the tumor samples (exclusive of normal samples). 95 snoRNAs were retained after filtering for read counts, were considered as continuous variables and were tested for their association with OS and RFS, using univariate Cox regression analysis, followed by permutation p-value. Twelve and ten snoRNAs were found to be significant for OS (Table 7.1) and RFS (Table 7.2), with permuted p-value < 0.1, which included all the snoRNAs identified from CC approach. Overall, thirteen snoRNAs were identified with prognostic relevance. Similar to the CC approach, risk scores were computed and the optimal cut-off points were

estimated to be -9.59 and -7.74 for OS and RFS, respectively, for dichotomizing the samples into low and high-risk groups. Risk scores were treated as dichotomous variables and were entered into univariate and multivariate Cox regression analyses for OS and RFS. For both OS (HR = 2.75, CI = 1.37 – 5.52, p-value = 0.005, Figure 7.3, Table 7.4) and RFS (HR = 2.42, CI = 1.33 – 4.42, p-value = 0.004, Figure 7.4, Table 7.4), risk scores were found to be independent prognostic factors, after adjusting for potential confounders. Patients belonging to high-risk group were associated with poor prognoses.

Figure 7.3 snoRNAs Kaplan-Meier plot for Overall Survival (case-only)

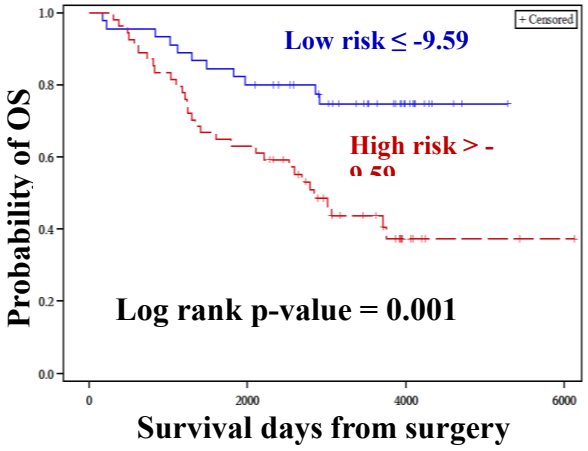
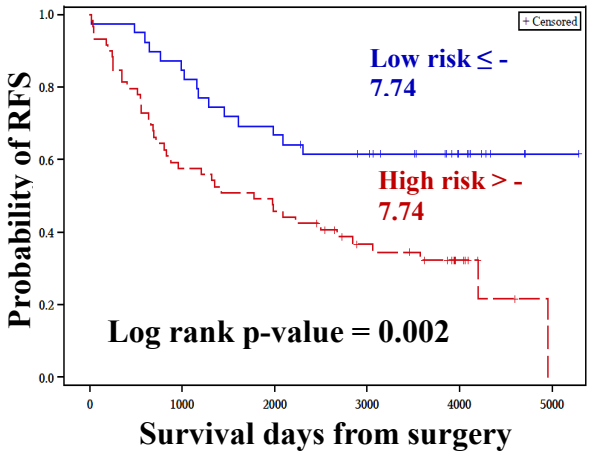


Figure 7.4 snoRNAs Kaplan-Meier plot for Recurrence Free Survival (case-only)



Figures 7.3 and 7.4: Kaplan-Meier plots for the risk score were constructed to determine the survival differences between low–risk and high–risk groups. Significant survival differences existed between the two risk groups, as indicated by the log–rank p–values. In both OS (Figure 7.3) and RFS (Figure 7.4), patients belonging to high–risk group showed poor prognoses.

Table 7.4 Univariate and multivariate results of snoRNAs for case–only approach

Parameter	Overall Survival				Recurrence Free Survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	P–value	HR (95% CI)	P–value	HR (95% CI)	P–value	HR (95% CI)	P–value
Risk score	2.95 (1.48–5.88)	0.002	2.75 (1.37–5.52)	0.005	2.44 (1.35–4.43)	0.003	2.42 (1.33–4.42)	0.004
Tumor stage	0.39 (0.2–0.78)	0.007			0.42 (0.22–0.78)	0.007		
Tumor grade	2.15 (1.06–4.39)	0.035	2.15 (1.04–4.42)	0.038	1.61 (0.91–2.86)	0.1		
Age at diagnosis	1.06 (1.02–1.09)	0.001	1.06 (1.02–1.09)	0.002	1.02 (0.99–1.05)	0.2		
TNBC status	0.93 (0.46–1.89)	0.83			0.76 (0.4–1.45)	0.41		

HR = Hazard ratio; CI = Confidence interval; TNBC = Triple negative breast cancer

Table 7.4: The two risk scores computed for CO (one for OS and one for RFS) approaches were found to be significant in the multivariate analysis ($p < 0.05$) after adjusting for potential confounders. For both the outcomes, patients with risk scores more than the estimated optimal cut–off points were associated with poor prognoses ($HR > 1$).

7.3.3 Concordance of findings between NGS and qRT-PCR

In NGS analysis, SNORD46 and SNORD89 were found to be down-regulated in tumors, relative to normal samples with a fold change of -7.38 and -4.07, respectively. When analyzed using qRT-PCR, these two snoRNAs showed the same direction of expression i.e., both the RNAs were down-regulated with $p < 0.05$, confirming the findings from NGS (Figure 7.5).

Figure 7.5 qRT-PCR confirmation of snoRNA expression

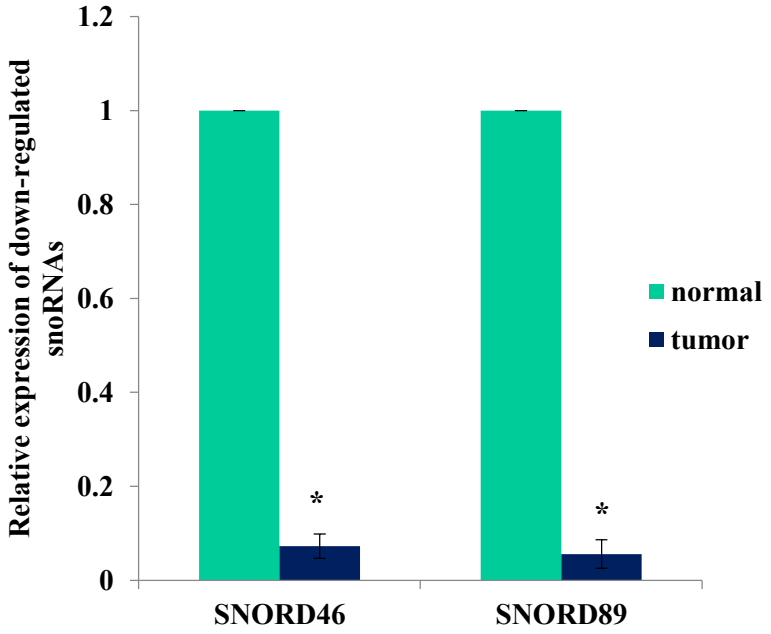


Figure 7.5: SNORD 46 and SNORD 89 were confirmed to be down–regulated in tumor, relative to normal samples using qRT-PCR platform. The Ct values obtained for snoRNAs were normalized to Ct values obtained for RNU6.

7.3.4 Insights into the regulatory functions of snoRNAs

Previous studies have reported that snoRNAs are often found in the intronic regions of protein-coding genes (snoRNA host genes) [18]. I also observed that out of 768 snoRNAs that were profiled in breast tissues (including normal and tumor tissues), 449 (i.e., > 50%) snoRNAs mapped to the intronic regions of protein coding genes (Appendix Table 9.8). It has also been suggested that snoRNAs can act as a source for other regulatory small non-coding RNAs, such as miRNAs [17,31,32] and piRNAs [39], implying a novel function and/or biological relevance for snoRNAs in gene regulation. In this study, the genomic coordinates of all the 768 snoRNAs were overlapped to that of mature miRNAs obtained from miRBase v 20. I observed that 8 mature miRNAs were contained

within six different snoRNAs. Further, the direction of fold change between these 8 miRNAs (data not shown) and snoRNAs were compared and was observed that 5 of these RNAs were expressed in the same direction in tumor tissues, relative to normal tissues (Appendix Table 9.9), hinting at the possibility that these miRNA-snoRNA pairs may be co–regulated.

I also extended this comparison to piRNAs and observed that 58 snoRNAs harbored piRNAs (Appendix Table 9.10). Of these, 35 piRNA-snoRNA pairs were expressed in the same direction in tumors, relative to normal tissues, i.e. if the piRNA was up-regulated in the tumor tissues; its corresponding host snoRNA was also up-regulated in the tumor tissues (Appendix Table 9.10). Additionally, from among the 35 pairs, 11 snoRNA-piRNA pairs showed DE with FC > 2.0 and FDR < 0.05 (Table 7.5) and I sought gene targets regulated by these snoRNA-piRNA pairs. In the in-house gene expression dataset, 628 genes were up-regulated and 2241 genes were down-regulated. Targets for the six down-regulated and five up-regulated piRNAs were interrogated using the 628 up-regulated and 2241 down-regulated genes, respectively. piRNA-mRNA targets with the specified criteria of alignment score and energy threshold score are summarized in Table 7.5. Gene ontology classifications of the genes identified as targets for piRNAs are summarized in Table 7.6.

Table 7.5 snoRNA-piRNA pairs with same direction of expression, fold change > 2.0 and FDR < 0.05

Host gene	snoRNA ID (Fold change)	Target RNA for snoRNA	piRNA embedded within snoRNA (Fold change)	mRNA targets for the embedded piRNAs
NOP56	SNORD110-201 (-24.22)	18S rRNA U1288	hsa_piR_019676 (-8.01)	DGKH,CLEC5A,ADAMDEC1,HOXC13,LRRC15,IQCH,WDR62
SNHG24	SNORD114-23-201 (-4.39)	unknown	hsa_piR_019102 (-6.43)	BPNT1,CASC5,KIF26B,PRAME,TLL2,ZC3H12D
SNX5	SNORD17-201 (-2.12)	28S rRNA U3797	hsa_piR_017033 (-2.17)	MAGEA4,PLGLB2,TNFSF4,FAM83D,CGA,FOSL1,GAS2L3,BRIP1,NCAPG,PLGLB2
HSPA9	SNORD63-201 (-3.80)	28S rRNA A4541	hsa_piR_000586 (-3.83)	None
AP1G1	SNORD71-201 (-2.09)	5.8S rRNA U14	hsa_piR_002158 (-2.78)	TPM3,DQX1
DDX39B, ATP6V1G2-DDX39B	SNORD84-201 (-2.24)	unknown	hsa_piR_001078 (-4.79)	GRM4,CENPI,CHRNA1,GPR26
TPT1	SNORA31-001 (1.58)	18S rRNA U218 and 28S rRNA U3713	hsa_piR_017184 (9.17)	TMEM47,TRPM3,ZNF462
PRRC2A	SNORA38-201 (14.44)	unknown	hsa_piR_004531 (54.1)	SLC34A1,SLC6A2,SEC31B,TFAP2C,TTC23,TXNIP,XPNPEP3,CNTN2
MRPL3	SNORA58-001 (8.65)	28S rRNA U3823	hsa_piR_020466 (4.06)	SLC27A1
SNHG16	SNORD1B-201 (5.08)	28S rRNA G4362	hsa_piR_018780 (18.45)	SMAD2,TNRC6B,TRIM9,UTRN,USP6,ANGPTL1,ARHGAP6,ALB,BCHE,CNTNAP3
CCAR1	SNORD98-201 (2.79)	18S rRNA G867	hsa_piR_000045 (4.85)	SFRP1,RSPO3,SPARCL1,WSCD1,ADRA2A,AVPR1A,ASPH,BCL6,CCDC25

Table 7.5: 11 snoRNA-piRNA pairs were observed to be present in the same direction with a fold change > 2.0 and FDR < 0.05. The host genes indicate the genes within which the snoRNAs are embedded. Since snoRNAs are involved in the modification of other RNAs, the target RNAs of the 11 snoRNAs have also been indicated. Also, since piRNAs are involved in gene regulation, the target mRNAs are listed corresponding to its piRNA.

Table 7.6 Gene ontology terms associated with genes targeted by piRNAs embedded within snoRNAs

Gene Ontology term	Genes involved	piRNAs targeting genes
Positive regulation of cell adhesion	ARHGAP6, CNTNAP3, CNTN2	hsa_piR_018780, hsa_piR_004531
Response to hormone	AVPR1A, BCHE, TXNIP	hsa_piR_000045, hsa_piR_018780 hsa_piR_004531
Regulation of apoptotic process	BCL6, ALB, SFRP1, TXNIP, FOSL1, GRM4, PRAME	hsa_piR_000045, hsa_piR_018780, hsa_piR_004531, hsa_piR_017033, hsa_piR_001078, hsa_piR_019102
Cell communication	BCL6, AVPR1A, BCHE	hsa_piR_000045, hsa_piR_018780
Regulation of cell proliferation	BCL6, SMAD2, ADRA2A, AVPR1A, ASPH, TXNIP, FOSL1, PRAME, TNFSF4	hsa_piR_000045, hsa_piR_018780, hsa_piR_004531, hsa_piR_017033, hsa_piR_019102
Transcription	BCL6, SMAD2, TXNIP, TFAP2C, ZNF462, BRIP1, FOSL1, HOXC13, TNFSF4	hsa_piR_000045, hsa_piR_018780, hsa_piR_004531, hsa_piR_017184, hsa_piR_017033, hsa_piR_019676
G-protein coupled receptor signaling pathway	GPR26, CENPI, DGKH, GRM4, CGA	hsa_piR_001078, hsa_piR_019676
Signal transduction	RSPO3, SMAD2, ADRA2A, ANGPTL1, AVPR1A, SFRP1, TXNIP, ADAMDEC1, GPR26, CENPI, DGKH, GRM4, CGA	hsa_piR_000045, hsa_piR_018780, hsa_piR_004531, hsa_piR_019676, hsa_piR_001078
Cell-cell signaling	SMAD2, SLC6A2, TFAP2C	hsa_piR_018780, hsa_piR_004531, hsa_piR_000045

7.4 Discussion

In this study, I identified 13 snoRNAs as potential novel prognostic markers for BC. Twelve snoRNAs were found to be associated with OS and ten snoRNAs were found to be associated with RFS, among which nine were common between OS and RFS for BC. We also explored their potential roles in gene regulation. snoRNAs are well known to be involved in post-transcriptional modification of other regulatory non-coding RNAs. Other

functions of snoRNAs, such as their association with various clinical factors or their involvement in gene regulation is also emerging.

The study design included two approaches (CC and CO) to identify the appropriate method for discovering prognostic markers. While the CC approach tests only the DE snoRNAs for association with outcomes [12,40], the CO approach is unbiased and interrogates all the snoRNAs retained after filtering, and is independent of the control tissues used [14,15]. Composite risk scores were calculated for the following reasons: (i) individual markers are not adequate to capture the complex interactions involved in conferring phenotypes and (ii) inclusion of all snoRNAs significant in the univariate analysis may contribute to data overfitting. The constructed risk scores were identified as potential independent prognostic factors for BC. Overall, in the CC approach, a total of 6 non-redundant snoRNAs were identified to be associated with disease outcomes (OS and RFS included), and as expected, a higher number of snoRNA markers (n = 13) were obtained from the CO approach, that included signatures from the CC approach as well. The same pattern, i.e., a higher number of markers in the CO approach (including those identified from the CC approach), was observed when I interrogated this dataset for miRNAs as prognostic markers [13]; highlighting the importance of considering CO approach for a wider inclusion of RNAs in a biomarker study.

To the best of my knowledge, this is the first study to report snoRNAs as prognostic markers for BC. In fact, none of the prognostic snoRNAs identified in this study have been reported in any of the other cancer types analysed thus far. These potentially novel prognostic markers for breast cancer need to be further independently validated to ascertain their role in BC prognostication. However, at this time, it is not certain if the 13

prognostic snoRNAs are specific to BC or if they share prognostic relevance in other cancer types. It is possible that with more genome wide studies focusing on understanding the clinical relevance of snoRNAs, we may be able to identify these snoRNAs in other cancer types. It would also be interesting to see if the identified snoRNAs show any subtype or tumor stage or grade specificity. In this pilot study conducted using 104 tumors; 62 samples belonged to Luminal A subtype (26 deaths and 37 recurrences) and 30 belonged to TNBC subtype (11 deaths and 13 recurrences). Given the current sample size and the number of events, it was not feasible to conduct further finer analysis based on stratified subtypes of BC.

I understand that a complex interplay exists between different classes of RNAs for normal developmental process and to maintain homeostasis. For instance, snoRNAs are known to be embedded within the intronic regions of protein-coding or non-protein coding genes. The well-studied function of snoRNA includes participation in post-transcriptional modifications of other RNAs such as rRNAs (involved in protein translation), snRNAs (involved in splicing mechanisms) and tRNAs (involved in protein translation). However, understanding of snoRNAs is slowly expanding towards gene regulation. snoRNAs have not been found to interact directly with mRNAs causing translational repression or mRNA degradation, similar to miRNAs. An alternative mechanism has been suggested, wherein the snoRNAs may get processed to form other regulatory RNAs such as miRNAs and piRNAs, well established regulators of gene expression. Figure 7.6 and Table 7.5 illustrate the complex interplay of these RNAs. In this regard, 450 snoRNAs were found to be embedded within the protein-coding genes (Appendix Table 9.8), and 6 miRNAs (Appendix Table 9.9) and 58 piRNAs (Appendix

Table 9.10) to be present within the genomic boundaries of snoRNAs. I also observed that the 11 snoRNA-piRNA pairs reported in Table 7.5 demonstrate same direction of alteration in tumor tissues, i.e., if the snoRNA was found to be down-regulated in the tumor tissues, its corresponding piRNA was also found to be down-regulated. We can speculate that some of the snoRNAs and piRNAs may be co-regulated and share a common promoter. However, the processing of these piRNAs/miRNAs from the snoRNAs needs to be ascertained, and further experiments are needed to understand their co-regulation, if any.

Figure 7.6 Complex interplay of snoRNAs with other RNAs

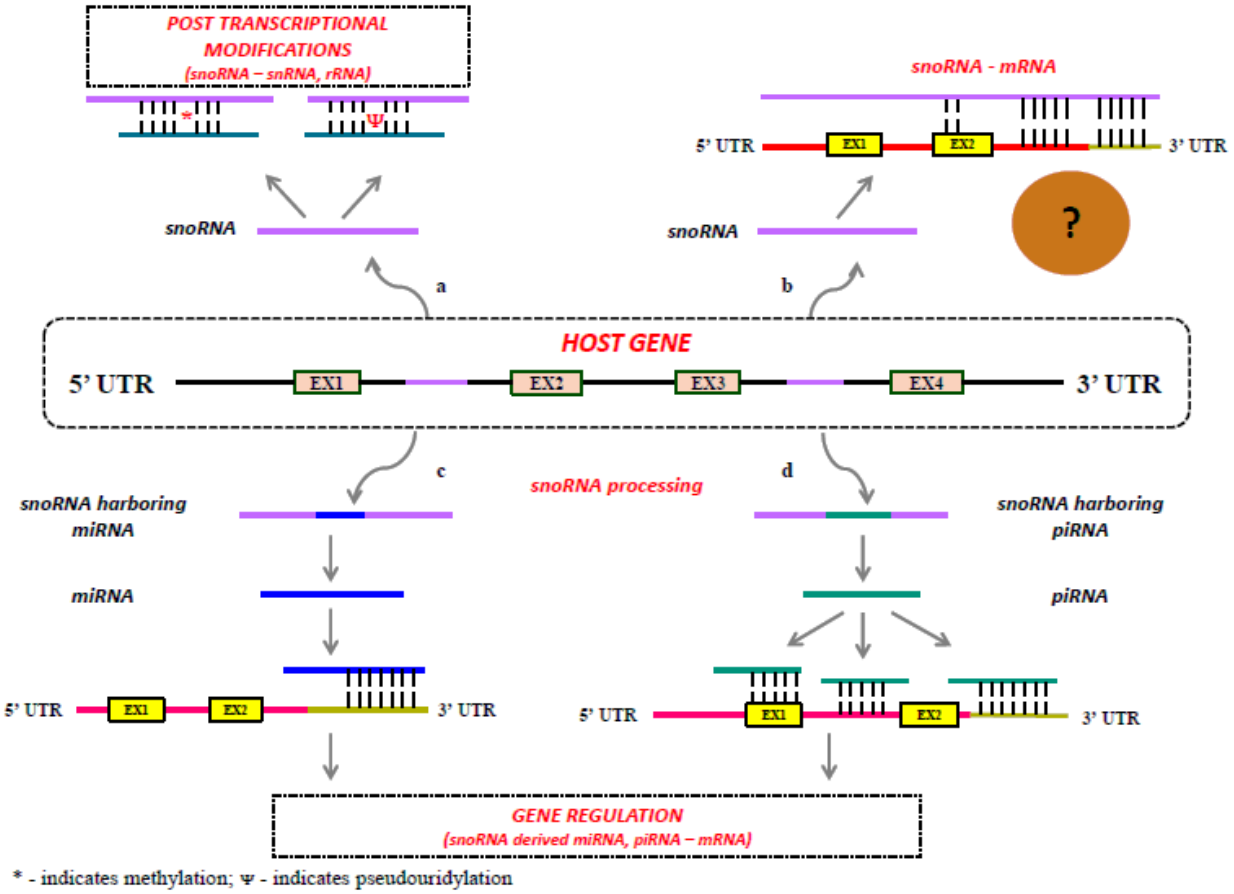


Figure 7.6: snoRNAs are involved in diverse biological functions. They arise from the intronic regions of protein coding / non-protein coding genes (host genes) EX represents exons. Black line indicates intronic regions and the purple line within the intronic regions indicate the snoRNAs. The canonical function of snoRNAs is its role in post-transcriptional modifications of snRNAs and rRNAs, which are involved in splicing mechanism and protein translation, respectively (a). One of the emerging roles of snoRNAs is its involvement in gene regulation. snoRNAs may act as a source for other small RNAs such as miRNAs (b, indicated in deep blue) and piRNAs (c, indicated in green). miRNAs and piRNAs are considered as master regulators of gene expression that may bind to the untranslated regions (3' UTR or 5' UTR), exons or introns and may promote either mRNA degradation or translation inhibition; implying the indirect role of snoRNAs in gene regulation. (d). The other unknown function of snoRNAs is its direct interaction with mRNAs through complementary base pairing. Till date, the direct interaction of snoRNAs with mRNAs has not been studied; however, this interaction might be a possibility based on the demonstrated subsets of snoRNAs embedding piRNAs and miRNAs, and their interactions with mRNAs through base pair complementarities; further research into this field may enhance our understanding on the direct role of snoRNAs in gene regulation.

Since these piRNAs originated from within the snoRNAs, the snoRNAs also shared certain degree of complementarity with the mRNAs (data not shownIt is not known if this degree of complementarity implies a direct interaction between snoRNAs and mRNAs and thus contributes to direct gene regulation. snoRNAs are larger in size (60–300 nt) compared to other regulatory small RNAs (miRNAs and piRNAs, 18–30nt). Therefore, the immediate challenge is to identify if canonical seed sequence motifs exist for snoRNAs to mediate gene silencing effects. However, at this point of time, we know that ectopic expressions of snoRNAs in a cell line or animal model could contribute to various cancer phenotypes such as cell proliferation, invasion and migration [25,41,42]. Interestingly, high expression of ACA11 was also found to contribute to increased resistance to chemotherapy in multiple myeloma [42], suggesting that snoRNAs may be important players for tumorigenesis. The targets identified for the 11 piRNAs (identified in this study) showed relevance in important tumorigenic pathways such as cell proliferation, cell adhesion and apoptosis (Table 7.6). Functional validation studies are thus warranted to confirm if these piRNAs interact directly with their corresponding targets to promote gene silencing.

7.5 Conclusions

In this study, I determined two aspects of snoRNAs: (i) comprehensive profiling of snoRNAs and their importance as prognostic markers for breast cancer and (ii) their possible roles in gene regulation. I report 13 (non-redundant) novel promising prognostic markers for breast cancer: 12 for OS and 10 for RFS. The contribution of snoRNAs to tumorigenesis is manifested through (i) their primary action in post-transcriptional modifications of other RNAs, and (ii) their processing to generate small RNAs that are

directly involved in gene regulation. While the first contribution of snoRNAs is well established, their role in gene regulation is only just emerging. Insights into these aspects could open up new avenues for the development of snoRNAs for diagnostic and therapeutic purposes.

7.6 References

1. Bogdanova N, Helbig S, Dörk T. Hereditary breast cancer: Ever more pieces to the polygenic puzzle. Hereditary Cancer in Clinical Practice 2013;11:1-16.

2. Hutchinson L. Breast cancer: Challenges, controversies, breakthroughs. Nat Rev Clin Oncol 2010;7:669-670.

3. Aparicio S, Mardis E. Tumor heterogeneity: Next-generation sequencing enhances the view from the pathologist's microscope. Genome Biol 2014;15:463-463.

4. Ribelles N, Perez-Villa L, Jerez JM, Pajares B, Vicioso L, Jimenez B, et al. Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. Breast Cancer Res 2013;15:R98-R98.

5. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-2826.

6. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 2004;116:843-854.

7. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 2004;116:855-862.

8. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res 2004;64:3087-3095.

9. Papagiannakopoulos T, Shapiro A, Kosik KS. MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. Cancer Res 2008;68:8164-8172.

10. Venkataraman S, Birks DK, Balakrishnan I, Alimova I, Harris PS, Patel PR, et al. MicroRNA 218 acts as a tumor suppressor by targeting multiple cancer phenotype-associated genes in medulloblastoma. J Biol Chem 2013;288:1918-1928.

11. Bertoli G, Cava C, Castiglioni I. MicroRNAs: New biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. Theranostics 2015;5:1122-1143.

12. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 2013;19:4477-4487.

13. Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics 2015;16:735-735.

14. Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

15. Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-microRNA signature. Gut 2010;59:579-585.

16. Rabinowits G, Gerçel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal MicroRNA: A diagnostic marker for lung cancer. Clin lung cancer 2009;10:42-46.

17. Martens-Uzunova E, Olvedy M, Jenster G. Beyond microRNA - novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett 2013;340:201-211.

18. Dieci G, Preti M, Montanini B. Minireview: Eukaryotic snoRNAs: A paradigm for gene expression flexibility. Genomics 2009;94:83-88.

19. Weinstein LB, Steitz JA. Guided tours: From precursor snoRNA to functional snoRNP. Curr Opin Cell Biol 1999;11:378-384.

20. Filipowicz W, Pogacic V. Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol 2002;14:319-327.

21. Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, et al. Prader-willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat Genet 2008;40:719-721.

22. Michel CI, Holley CL, Scruggs BS, Sidhu R, Brookheart RT, Listenberger LL, et al. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. Cell metabolism 2011;14:33-44.

23. Ronchetti D, Mosca L, Cutrona G, Tuana G, Gentile M, Fabris S, et al. Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia. BMC Medical Genomics 2013;6:1-11.

24. Xu G, Yang F, Ding CL, Zhao LJ, Ren H, Zhao P, et al. Small nucleolar RNA 113-1 suppresses tumorigenesis in hepatocellular carcinoma. Molecular cancer 2014;13:216.

25. Okugawa Y, Toiyama Y, Toden S, Mitoma H, Nagasaka T, Tanaka K, et al. Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. Gut 2015.

26. Ravo M, Cordella A, Rinaldi A, Bruno G, Alexandrova E, Saggese P, et al. Small non-coding RNA deregulation in endometrial carcinogenesis. Oncotarget 2015;6:4677-4691.

27. Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkersfor non-small-cell lung cancer. Molecular Cancer 2010;9:198-207.

28. Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, et al. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. Int J Cancer 2015;136:E623-E629.

29. Dong X, Guo P, Boyd J, Sun X, Li Q, Zhou W, et al. Implication of snoRNA U50 in human breast cancer. Journal of Genetics and Genomics 2009;36:447-454.

30. Su H, Xu T, Ganapathy S, Shadfan M, Long M, Huang TH, et al. Elevated snoRNA biogenesis is essential in breast cancer. Oncogene 2014;33:1348-1358.

31. Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, et al. Article: A human snoRNA with MicroRNA-like functions. Mol Cell 2008;32:519-528.

32. Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res 2011;39:675-86.

33. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 2001;25:402-408.

34. Germain DR, Graham K, Glubrecht DD, Hugh JC, Mackey JR, Godbout R. DEAD box 1: A novel and independent prognostic marker for early recurrence in breast cancer. Breast Cancer Res Treat 2011;127:53-63.

35. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl genomes 2016: More genomes, more complexity. Nucleic Acids Res 2016;44:D574-D580.

36. Sai Lakshmi S, Agrawal S. piRNABank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173-7.

37. Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-10.

38. Hall JS, Taylor J, Valentine HR, Irlam JJ, Eustace A, Hoskin PJ, et al. Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. Br J Cancer 2012;107:684-694.

39. Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, et al. A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. Nucleic Acids Res 2015;43:10474-10491.

40. Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-1184.

41. Mei Y, Liao J, Shen J, Yu L, Liu B, Liu L, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. Oncogene 2012;31:2794-2804.

42. Chu L, Su MY, Maggi L,Jr, Lu L, Mullins C, Crosby S, et al. Multiple myeloma-associated chromosomal translocation activates orphan snoRNA ACA11 to suppress oxidative stress. J Clin Invest 2012;122:2793-2806.

8 Discussion, conclusions and future work

Breast cancer (BC) is a complex polygenic disease that exhibits histological and molecular heterogeneity. Optimal management of BC has been a challenge since patients exhibiting similar clinical characteristics respond differently to the same guideline based therapies, signifying inter-individual variations in outcomes. It is necessary to identify biomarkers with higher specificity and sensitivity than the currently used hormone receptor expressions to guide treatment decisions. Primary objective of this thesis was to evaluate the clinical relevance of sncRNAs for BC. I conducted this study with well-annotated clinical specimens having long years of follow-up and using the state of the art technology – NGS. By far, genome-wide profiling of sncRNAs using NGS has been limited, or absent for some classes of sncRNAs and this study complements such efforts towards a comprehensive search for biomarkers. Four classes of sncRNAs were interrogated for their association with breast cancer prognosis in the same breast cancer tissue sample sets (outcomes of interest: overall survival and recurrence free survival) – miRNAs, piRNAs, tRNAs and snoRNAs. Each of these RNAs has defined roles in biology and yet one of the characteristics that are common between these RNA classes is their ability to regulate gene expression at post-transcriptional level (piRNAs may also function in gene regulation at a transcriptional level in somatic cells) and miRNA mediated gene regulation is well explored [1,2]. Deeper exploration of piRNAs has identified their potential usefulness as master regulators of gene expression [3,4]. The contribution of the latter two classes of sncRNAs to gene regulation is slowly beginning to be understood in an indirect fashion [5-9]. Since the clinical significance of miRNAs in breast cancer is well known (explained in earlier chapters – 1.2.1 and chapter 4), I began

my study by identifying miRNAs as prognostic markers for BC. Since profiling study is genome-wide and comprehensive in data mining approaches, I was able to reproduce the findings from others and in the process also identified two novel miRNAs for BC prognosis. I was also able to replicate the study findings in an external dataset. Reproducibility of previous findings and replication of the study findings in an independent/external dataset added strength and confidence to the study findings. These promising results from miRNAs gave me the confidence to interrogate the other three sncRNA classes.

The following inferences were made from this work:

1. miRNAs: Eleven miRNAs (hsa-miR-210-3p, hsa-miR-15a-5p, hsa-miR-660-5p, hsa-miR-146b-5p, hsa-miR-374a-3p, hsa-miR-374a-5p, hsa-miR-27a-3p, hsa-miR-574-3p, hsa-miR-221-3p, hsa-miR-196a-5p, hsa-miR-425-5p) were found to be associated with OS and four miRNAs (hsa-miR-210-3p, hsa-miR-425-5p , hsa-miR-193b-3p , hsa-miR-15a-5p) were found to be associated with RFS. Gene targets identified for the prognostically significant miRNAs were found to be involved in breast tumorigenesis through several key processes such as cell migration and proliferation, angiogenesis etc.,. The prognostic significance of ten of the above mentioned miRNAs from published literature were also confirmed in this study. In addition, NGS profiling helped identify two novel miRNAs (miR-574-3p and miR-660-5p) for breast cancer prognosis.
2. piRNAs: Four piRNAs (hsa_piR_009051, hsa_piR_021032, hsa_piR_015249, hsa_piR_020541) and ten piRNAs (hsa_piR_017061, hsa_piR_009051, hsa_piR_021032, hsa_piR_004153, hsa_piR_017716, hsa_piR_019914) were

identified to be significant for OS and RFS, respectively. None of these piRNAs have been reported in BC literature for their association with prognosis.

a. The prognostic significance of mRNAs coding for PIWI proteins, which are crucial players in piRNA biogenesis, was also demonstrated for the first time in breast cancer literature.

b. Both upstream (PIWI mRNAs) and downstream (mRNA gene targets) partners for piRNAs were dysregulated, indicating an overall perturbation in the piRNA pathway in a cancer setting. Several targets were identified for the prognostic piRNA (or subsets) and Gene Ontology terms were enriched for the biological and molecular processes involved in the development of cancer.

3. tRNAs: Fourteen tRNAs each, were found to be associated with **OS** (chr6.tRNA166-AlaAGC, chr17.tRNA10-GlyTCC, chr6.tRNA147-SerAGA, chr6.tRNA145-SerAGA, chr6.tRNA5-SerAGA, chr16.tRNA2-ArgCCT, chr6.tRNA50-SerAGA, chr12.tRNA8-AlaTGC, chr6.tRNA148-SerTGA, chr6.tRNA172-SerTGA, chr6.tRNA143-LysTTT, chr14.tRNA2-LeuTAG, chr6.tRNA51-SerTGA, chr9.tRNA4-ArgTCT) and **RFS** (chr6.tRNA166-AlaAGC, chr1.tRNA80-GluCTC, chr1.tRNA77-GluCTC, chr6.tRNA87-GluCTC, chr1.tRNA74-GluCTC, chr1.tRNA71-GluCTC, chr1.tRNA59-GluCTC, chr6.tRNA77-GluCTC, chr1.tRNA118-HisGTG, chr6.tRNA152-ValCAC, chr1.tRNA116-GluCTC, chr2.tRNA19-GlyGCC, chr6.tRNA128-GlyGCC, chr1.tRNA133-GlyCCC), respectively. This is the first time in literature that the prognostic relevance of tRNAs were interrogated for.

4. snoRNAs: The prognostic significance of twelve snoRNAs for

   OS (SNORA7A, SNORD94, SNORD92, SNORD105, SNORA7B, SNORD85, SNORD46, SNORD100, SNORD104, SNORD84, SNORD82, SNORD14E)

   and ten snoRNAs for

   RFS (SNORD92, SNORA7B, SNORD14E, SNORD46, SNORD105, SNORD94, SNORD100, SNORD89, SNORD82, SNORA7A) were demonstrated for the first time, for breast cancer. snoRNAs were also observed to harbor other regulatory small non-coding RNAs such as miRNAs and piRNAs.

5. The prognostic significance of miRNAs, piRNAs and tRNAs associated with OS were validated in an external dataset obtained from TCGA.

8.1 Challenges

This is one of the studies which has utilized a large sample set for BC ductal carcinomas with complete clinical annotation and has looked at the family of sncRNAs from a holistic view. I was fully aware that undertaking a comprehensive, high throughput genomics study for prognostic markers is never free from challenges and this thesis was no exception. Following are some of the challenges that were addressed in this study:

8.1.1 Choice of normal tissues

More often than not, studies conducted so far have considered the tissues adjacent to tumors as relatively normal tissues for use as reference tissues [10,11]. One of the reasons for the choice of these tissues may be attributed to the ease in procuring these samples [12], when compared to the breast tissues obtained from an apparently healthy individual (reduction mammoplasty). Although these normal tissues are scored for the presence of

any tumor cells and only those which do not have any tumor cell infiltration are considered, still there are reports to show that these relatively normal tissues are often undergoing the process of malignant transformation and may not be the most appropriate choice for use as normal tissues [13]. Another challenge with the use of tumor adjacent normal tissues is the distance used for determining the tumor adjacent normal tissues. So far, there has not been a standard or established cut-off for the distance. Sometimes, adjacent normal tissues may not be evident up to a distance of 7 cm away from the main tumor [14] or the resection margin might expand to more than 10 cm [13,15].

Studies have been conducted to understand the molecular differences between the tumor adjacent normal tissues, normal tissues obtained from healthy individuals and tumor tissues. For instance, in a study conducted by Sanz-Pamplona et al., mRNA levels from 98-paired adjacent normal mucosa were compared with colorectal cancer tissues and with 50 colon mucosa from healthy donors [13]. It was observed that a number of genes were activated in the adjacent mucosa, similar to the tumor tissues and the activation of these genes was not observed in the normal tissues from healthy donors. In fact, principal component analysis of the three tissue sets identified different clusters for the three tissue types, suggesting that considerable differences existed between the adjacent normal tissues and healthy normal tissues. Further, a total of 895 genes were found to be differentially expressed between the two normal tissues. Functional enrichment analysis and network analysis of the activated genes from adjacent normal tissues identified pathways significant for tumorigenesis, implying that the adjacent normal tissues may mimic the effects of tumorigenesis. The same was validated in publically available datasets, consistent with the view that adjacent stromal cells cross-talk with tumor cells

within this micro environment (called as “the field effect”). Therefore use of adjacent normal tissues as a reference may result in misleading interpretations. The effect on the adjacent normal tissues has also been echoed in prostate tissues as well [16]. In this study by Chandran et al., prostate tumor tissues and tumor adjacent normal prostate tissues were compared against the tissues obtained from apparently healthy individuals. Both the tumor tissues and adjacent normal tissues exhibited similar mRNA expression patterns when compared to the other normal tissues. Even in breast cancer studies for gene (mRNAs) expression, normal tissues obtained from healthy individuals from mammoplasty have been identified as appropriate for use as a reference than the adjacent normal tissues. Several gene expression differences were noted between the healthy normal tissues and adjacent normal tissues [17,18]. Similar anomalous patterns of expression were reported even for non-coding RNAs when tumor adjacent normal vs. reduction mammoplasty normal tissues were compared. In support of this general premise, differential miRNA expressions were reported between reduction mammoplasty and tumor adjacent normal tissues [19]. Differentially expressed long non-coding RNAs were absent when tumor adjacent normal tissues vs. tumor tissues were compared. However, differentially expressed transcripts were identified when tumor tissues were compared with reduction mammoplasty specimens. It is safe to infer from the above, that reduction mammoplasty specimens may be the ideal tissue source to serve as a reference [20]. Breast tissues obtained from women undergoing reduction mammoplasty surgery are free of cancer at the time of surgery; a pathological assessment is also mandatory to confirm absence of tumor cells. The same was confirmed before conducting this study.

8.1.2 Choice of tissue preservation techniques

There are two widely adopted methods for long term preserving of tissue specimens – fresh frozen or flash frozen (FF) and formalin fixed paraffin embedded tissues (FFPE). While FF technique offers the advantage of preserving the biomolecules intact [21], it has not been cost effective and preserving such tissues is cumbersome when small tumors are detected. On the contrary, FFPE tissue blocks offer advantages for clinical studies because of the ease of obtaining the blocks from archival specimens with long follow-up periods and offer possibilities for translation research [21]. FFPE blocks can be stored at room temperatures for many years and are easier to handle, thereby are more cost effective [22]. However, extracting nucleic acids from FFPE tissue blocks has been challenging due to the extensive cross-linkages of macromolecules (such as RNA, DNA and proteins) by formaldehyde containing fixatives [23,24]. However, it has been postulated that the problem of cross-linking should not deter the efforts in profiling small RNAs of <200 nt and in particular mature miRNAs (typically 21 nucleotides in length) for the following reasons [21]: (i) their slower degradation rate [25], (ii) smaller size and (iii) lack of poly A tails [26]. Several published studies to-date have demonstrated the feasibility of profiling small RNAs from FFPE tissue blocks. Since FF tissues are considered to preserve RNA integrity, often expression profiles of small RNAs obtained from FFPE tissues are compared with FF tissues. The exploration of profiling miRNAs from FFPE tissues started with one of the early studies conducted by Xi et al. excellent correlation (correlation coefficient between 0.86 and 0.89) was observed between FF and FFPE tissue specimens [26]. Another important observation was that miRNAs remained stable even in a ten year old or 28 year old sample preserved as FFPE tissue block [26,27].

Following this study, several other studies were published, confirming Xi et al's study findings in different tissues [21,22,28,29]. Therefore, with these precedents, I used both FF and FFPE tissue blocks.

8.1.3 Choice of RNA extraction protocol

This has been explained in chapter 2 (section 2.3).

8.1.4 Choice of profiling platform for small RNAs

miRNAs are predominantly profiled using any one of the three profiling platforms – microarray, qRT-PCR or NGS. Each of these profiling platforms has its own advantages and disadvantages in profiling, with NGS having an edge over the other two methods [30-32]. This has been elaborated in 1.3.3.

In this regard, some of the studies have focused on platform-based comparison of miRNAs and other small RNAs from FF and FFPE tissues. One fine example is the study conducted by Meng et al [33]. Tissues from six different cancer types were subjected to small RNA sequencing. Several important observations were recorded: (i) the stability of miRNAs were not affected by long storage time (their window was nine years); (ii) a comparison of highly degraded sample (as assessed by RNA integrity number) yielded similar number of aligned reads and miRNAs; (iii) the correlation between FF and FFPE samples showed good concordance (at least 85%) and, (iv) expression profiles of other sncRNAs (tRNAs and snoRNAs) was also shown to be feasible from FFPE tissues. Hence, NGS was chosen as a profiling platform for my work.

8.1.5 Challenges in data analysis are discussed below.

8.1.5.1 *Tackling batch effects*

Batch effects are non-biological sources of variations that may give rise to spurious results. Article by Leek et al., clearly explains the causes, effects of, methods to detect and correct for non-biological sources of variation [34]. Batch effects may arise because of different processing dates of samples, handling of samples by different individuals, different reagents etc. Although it is not possible to contain all the sources of errors, it is important that we handle these issues and correct the dataset accordingly. Ideally, the best way to avoid batch effects is to run all the experiments on the same day under identical experimental conditions and by the same person. However, this may not be feasible at all times and the current study is no exception. The samples in this study were sequenced in four different batches and in only one of the batches all of the normal samples were included. Reasons for processing the samples in different batches are explained in 2.5.1. One of the principal ways to determine batch effects is through Principal Component Analysis (PCA). In the presence of batch effects, sample clustering would not be based on biological reasons such as normal and tumor samples but would be based on the dates or batches in which the samples were processed. Another way to identify batch effects is through unsupervised hierarchical clustering. Partek Genomics suite offers one more option to quantify the amount of variation in the data using ANOVA model. Other software tools may help in assessing the presence of batch effects but may not have the option to adjust the dataset for batch effects. However, PGS offers this and for this study, I assessed the amount of variation arising from technical sources before and after batch effects, using PCA plots and ANOVA. PCA calculates the

variance between different batches and clusters the samples according to the amount of variance present. Presence of batch effects will cluster the samples according to the batches in which they were sequenced and will not cluster based on the actual biological differences such as tumor and normal samples. Hence, PCA may aid in visual representation of the presence/absence of batch effects in a dataset, while ANOVA model helps in quantifying the amount of variance existing because of batch effects. Often, studies using NGS do not mention about batch effects. One of the reasons for this may be because not many studies have used a large data set; thereby reducing the non-biological sources of variation and minimizing the need to correct for batch effects.

8.1.5.2 *Dimensionality reduction*

Data generated from high throughput techniques suffer from the problem of dimensionality, where the numbers of markers are much higher than the samples used. The real challenge in such datasets is to distinguish signal from noise. One way to address this issue is to set a filtering cut-off based on the expression level of RNAs profiled. Till date, there has not been a specified threshold or a logical reason for the use of a specific filter criterion. When investigators map the reads to different databases, the output generated will have all the RNAs with at least one read count. It is possible that these RNAs are expressed in only one of the samples used or in very few of the samples. It is best to get rid of these molecules as their contribution to real biological inference will be minimum. Therefore some studies employ a cut-off to include all the RNAs with a total read count of 3 or more [35-39]. Till date, the cut-off is not yet standardized and it is up to the discretion of the authors. Although this cut-off takes into account the expression of RNAs in the overall data, yet it does not consider the frequency of expression in the

samples, which I believed was important to increase the chance of reproducibility. Higher the expression and the frequency of occurrence of a RNA, higher the probability of the RNA being expressed in other datasets. With this premise, I imposed a cut-off of a minimum of ten read counts [39] in at least 90% of the samples. This reduced the number of markers for further investigation to a great extent. As a next step, since the objective of my thesis was to identify prognostic markers, all the sncRNAs (DE list in case of case-control approach and all the filtered RNAs in case of case-only approach) were subjected to univariate Cox proportional hazards regression model. To further reduce the number of data points or variables (in this case, the sncRNAs) and to add confidence to the univariate results, a permutation test was run for the Cox model and all the sncRNAs with $p \leq 0.1$ were considered for multivariate analysis. Since permutation test is feasible only for continuous variables, for the first pass of tests, sncRNAs were considered as continuous variables.

8.1.5.3 *Collinearity*

Quite often than not, high throughput data also suffers from the problem of collinearity (correlation between two markers). The problem of collinearity is more pronounced when we build multivariate models for Cox analysis, in which the inclusion of collinear variables may lead to the generation of instable coefficients [40]. In such cases, the multivariate model may over estimate or under estimate the associations and may not yield reliable results. Another important factor for consideration is the number of variables that can be included in a model. For a stable model, the rule of thumb in statistics is that every variable that is entered for multivariate analysis should correspond to at least ten events [41]. For instance, if I enter 4 variables in the model, I should have at

least 40 events (e.g. deaths or recurrences) in my sample set. With all this in mind, a risk score was constructed using all the sncRNAs significant in the permutation test. The advantage of using a risk score is that it is not affected by the problem of collinearity and also helps in reducing the number of variables to be entered into the multivariate model. Another advantage of using a risk score is that it takes into account, the effect of all the molecules which correlates to the fact that biology is not driven by a single molecule but by interaction of molecules.

8.1.6 External validation

The confidence in a biomarker study largely depends on its reproducibility in external datasets with large sample sizes. This adds strength to the initial study findings and may eventually lead to generalizing the initial study findings. However, it is not an easy task to obtain an external dataset because several factors need to be considered before choosing an external dataset. These include (i) number of samples available, (ii) histological and molecular subtypes of breast cancer used, (iii) profiling platform used for the study, (iv) number of events, (v) the amount of clinical information available, (vi) number of years of follow-up and (vii) the population used for the study. Since this study was performed using a sequencing platform, I was searching for a dataset that has used sequencing platforms. To the best of my knowledge, only one study has profiled large number of BC samples using Illumina Genome analyzer. However, this study had clinical information for TNBC and HER2+ samples alone and not for Luminal samples [42]. Since Luminal samples formed a large portion of the dataset that I used, I couldn't consider this dataset for validating my findings. Although I could have considered studies which had used other platforms such as qRT-PCR or microarray, it would not have been possible to

validate piRNAs, snoRNAs and tRNAs. As such large scale profiling has not been attempted for these three RNAs. Thus identifying an ideal external dataset was an arduous task. An independent dataset could have been generated from within the same population but a biomarker study is considered to be more robust if the validation is carried out in datasets that has samples from different geographical locations. With this in mind, I probed into the dataset generated by TCGA, which had about 1,088 BC samples. TCGA is a large project that collects samples from different geographical locations, mostly within United States of America, and generates a wide variety of data, including small RNA sequencing data. I have previously explained the different filtering criteria to choose the samples from TCGA (section 2.2.2). As mentioned, TCGA samples were sequenced in two different platforms that generate different volumes of output. Although samples sequenced using Illumina Genome analyzer would have been more appropriate for validation, the number of events in this dataset was insufficient to conduct a survival analysis; leaving samples from Illumina HiSeq for validation. Several differences existed between the discovery and validation cohorts: (i) platform differences, (ii) TCGA samples were preserved as FF tissues, (iii) information on tumor grade was missing and (v) percent tumor cellularity was relatively less when compared to the discovery cohort. Despite these differences, TCGA dataset served as the best resource available for validation. However, due to these differences between the two cohorts, the same expression values for sncRNA classes could not be obtained and hence the same risk score cut-offs could not be validated. Nevertheless, the new risk scores showed the same trend and direction of effect and in the case of piRNAs, the risk score even showed statistical significance with $p < 0.05$ in the multivariate analysis, similar to the discovery

cohort. A limitation with this dataset was that the signatures for recurrence free survival could not be tested in TCGA dataset, as the information on recurrence was minimal.

8.2 Strengths of this study

- All the samples have complete clinical annotation, including information on outcomes and have a median follow-up period of 8 years (6 months to 16.78 years).
- The choice of reduction mammoplasty specimens in place of adjacent normal tissues is an added strength to this work for reasons explained in 8.1.1.
- Use of NGS platform for profiling sncRNAs and identifying prognostic markers (see 1.3.3 and 8.1.4).
- Analysis of NGS generated datasets including quantification and adjustment of the data for batch effects. This has helped increase the confidence in the study findings.
- Used two statistical approaches to identify prognostic markers: Case-control and case-only. Adopting either of the two approaches is a common sight in literature, whereas use of both the approaches in the same study is rare. Based on the findings from all the sncRNAs interrogated in this thesis, adopting a case-only approach may be more appropriate as it enables a comprehensive capture of prognostic markers, which also includes markers identified from the case-control approach. Case-only approach may be an unbiased study design as it is not influenced by any relative expression differences from the control tissues. In this study, prognostic significance of individual markers were not investigated; instead

a composite risk score was constructed to understand the collective contribution of a molecular signature to clinical outcomes.

- miRNAs are well established prognostic markers.. The current study is an unbiased study, i.e., it is not focused on candidate miRNAs, but instead has attempted a comprehensive profiling using NGS. Total of 12 miRNAs were classified as potential prognostic markers in this study (chapter 4), ten of which have already been reported by others (not in a single study or by the same group in independent attempts, but represent the collective wisdom over several years). We report the same direction of effect in this study for all of reported miRNAs in a single study, lending credence to our study design, which is important for these markers to make inroads into BC prognostication.
- Apart from identifying literature reported markers, two miRNAs were identified as novel prognostic markers for BC.
- Clinical significance of the identified sncRNAs reported here is unknown until now; piRNAs, tRNAs and snoRNAs have shown promise as potential independent prognostic factors for BC.
- Validating the findings from a discovery cohort in an external dataset is considered crucial for a biomarker study. Since external validation datasets are generally obtained from a different geographical location, validating in an external dataset strengthens the initial study findings ruling out potential bias in case recruitment, geographical location, platform differences in NGS, handling of specimens (pre-analytic variables) etc. The prognostic significance of miRNAs,

piRNAs and tRNAs associated with overall survival were validated in dataset obtained from TCGA.

- Although the focus of this study was to identify prognostic markers, I also sought to gain insights into the biological relevance of some, if not all of the markers identified. miRNAs and piRNAs (of late) are being known as global regulators of gene expression. Several target prediction software exist such as TargetScan, miRanda etc., but these do not predict targets based on specific tissues or context and are therefore innumerous. One of the common ways to address this is to identify targets which are common to two or more databases and subsequently perform functional studies to confirm their interaction. In any case, we may still not be able to observe if the targets are indeed expressed in human tissues. Therefore, I used a breast tissue gene expression dataset to identify mRNA targets. The advantage of using such a dataset is that (i) we will only identify targets which are expressed in tissues, and (ii) it may serve as proxy for functional validation of the identified targets. This also reduces the need for interrogating two or more databases, each of which has its own advantages and disadvantages. For instance, I used TargetScan to predict targets for miRNAs and then overlapped with genes identified from breast tissue dataset as this database is more common and is regularly updated (latest version was released in June 2016). However, TargetScan may not be ideal for other RNAs such as piRNAs as it is not possible to interrogate targets from a tissue dataset and as such, no tool exists for predicting piRNAs. Therefore, I used miRanda as an interface to identify piRNA targets from gene expression tissue dataset. As such, miRanda algorithm

also does not predict targets for piRNAs but we can download the algorithm and use it as an interface to identify targets from human tissue datasets.

- The study design also helped to identify miRNAs and piRNAs embedded within snoRNAs and tRNAs, thus identifying newer function for these molecules.

8.3 Future work

8.3.1 Validation of the study findings

Overall, several important conclusions (outlined above) have been made from this study. Even though miRNAs are fairly well established as prognostic markers, identification of two novel miRNAs for BC prognosis indicates that the potential of miRNAs has not been tapped completely. For instance, 3,707 novel mature miRNAs have been identified in the year 2015 by Londin et al [43]. It would be interesting to see if these novel miRNAs are expressed in breast tissues, understand the various functions of these molecules and to see if these can also perform as prognostic markers for BC. With constant improvements in technological platforms and analytical approaches, newer molecules are being identified; we are still learning different ways to understand the dataset and to identify the most appropriate markers. Real test for this continuous learning lies in validating the identified signatures in other datasets. In this regard, we have seen some positive results from TCGA dataset even though the sample size and the number of events were low. However, sncRNAs associated with RFS have not been confirmed using any other dataset. Both OS and RFS signatures therefore need further independent validations. The datasets to be generated should pay attention to sequencing platforms for profiling of sncRNAs particularly for miRNA, piRNA, snoRNA and tRNA capture to enable replication of prognostic findings reported in this thesis. In an ideal

scenario of biomarker study validation, one would expect that the same risk score cut-off be validated in the new dataset as well. However, with the developments made in sequencing technologies, it may not be possible to sequence the samples using the same platform that was used for the discovery cohort. So far, it has not been possible to combine datasets generated from two different sequencing platforms. Lack of knowledge and consensus in the scientific community on methods to normalize datasets from diverse sequencing platforms is a current limitation. However, I believe that in the coming years, this would be made possible, in which case, validating the same risk score cut-off would be feasible. In cases where it may not be possible (studies outlines here), constructing a separate risk score using the same RNAs and estimating a new cut-off would be the best choice (as described in this thesis). One could argue that adopting such a method would make the identified risk score as data dependent and may not be appropriate in a clinical setting. But on a different perspective, this could also be considered as a completely unbiased method which will add more strength to the findings and enable us to perform further experiments with more confidence. We should also remember that the science that is evolving around NGS and biomarkers is new and until we establish gold standard protocols, we have to explore different methodologies to arrive at the best possible solution.

8.3.2 Subtype specific markers

It is always interesting to find out if the identified markers are specific for any particular subtype of breast cancer. This could not be attempted in the current thesis, mainly due to sample size. Once the identified signatures are independently validated (with due attention to the sequencing platform and batch effects), a meta-analysis could

be attempted for specific subtypes of BC or based on tumor-stage to help us understand if these sncRNAs can act as early stage markers. Understanding the survival patterns early in the disease trajectory may help modify the treatment regimen and improve the survival rate of patients.

8.3.3 Novel markers vs. existing markers

The purpose of attempting a biomarker study is to identify novel prognostic or predictive markers for use as stand-alone or in combination with traditional markers to improve the specificity and selectivity. One needs to attempt an analysis of combinations of sncRNA markers (independent of ER, PR and Her2) to assess whether these markers perform better in a clinical setting. It is relatively easy to develop point-of-care diagnostics based on sncRNA signatures. As a first step, robust platforms (NanoString Technologies) are now available; several laboratories are developing diagnostics and prognostics based on NanoString including my parent department (Laboratory Medicine and Pathology, University of Alberta).

8.3.4 Markers from circulation

In general, tumor based markers are considered to be invasive. Even though our basic understanding on markers stems from analyzing tumor specimens, a more useful application of tumor based findings would be to expand the study to easily procurable and minimally invasive samples such as serum or plasma. The source of sncRNAs in circulation is debatable. Nevertheless, the general notion is that the expression level of sncRNAs in circulation is relatively less. Since the RNAs chosen in this thesis are considered to be highly expressed, it may be possible to identify the same set of sncRNAs in these less invasive samples. However the logistics involved in analysing serum/plasma

samples are manifold and may include the time of sample collection, profiling of sncRNAs, especially the low expressors, among other factors. Therefore a careful study design is utmost important before conducting such a study.

8.3.5 Functional characterization

New therapeutics developments are possible if we understand the biology of the markers and their contribution to breast tumorigenesis. In this respect, an insight into the functional significance of the identified markers, especially the miRNAs and piRNAs were gained. Since the potential gene targets were identified from breast tissue gene expression dataset, it served as a proxy for functional validation of the targets, following in silico prediction. Further, gene functions of the identified targets were understood by performing a functional enrichment analysis. Nevertheless, since I have not performed any cell based assays to demonstrate their exact mechanism, I could not delineate the exact contribution of these RNAs in cellular functions. As such, not all predicted gene targets have been functionally validated and functional characterization of some of these sncRNAs would be a significant contribution to the sncRNA literature. As a starting point, several sncRNAs identified from this thesis work could be of immediate interest:

a. The two novel miRNAs identified: miR-574-3p and miR-660-5p. Both these miRNAs have displayed a fold change more than 2.0 and the corresponding gene targets have already been identified. The challenge however is to narrow down the target genes (mRNAs) for validation using luciferase constructs. The functional significance of both these RNAs remains unknown for breast cancer. In the case of miR-574-3p, gain-of-function analysis has revealed inhibition of cell proliferation, migration and invasion in bladder cancer cell lines and

MESDC1 gene was found to be a direct target for miR-574-3p [44]. miR-574-3p is also known to modulate tamoxifen resistance in breast cancer [45], an area of immediate interest in the miRNA field and BC. On the other hand, there has not been a single study that has demonstrated the functional significance of miR-660-5p in BC or in any other cancer.

b. In this study, the gene targets identified for piRNAs were based on a specific alignment score and energy threshold score, which are indicative of the strength of piRNA-mRNA interaction. However, one of the piRNAs that caught my attention for immediate characterization was hsa_piR_021032. This piRNA exhibited a fold change of 3.61 in tumors, relative to normal tissues but the unique feature of this RNA is that it is predicted to target one of the PIWI genes – PIWIL2. I observed an inverse relationship between the expressions of hsa_piR_021032 and PIWIL2. This sort of feedback mechanism is novel in literature and requires further confirmation. Such an understanding would help us appreciate the piRNA pathway better and may open up avenues for therapeutic explorations using these molecules.

c. We are beginning to understand that sncRNA molecules such as tRNAs and snoRNAs may contribute to gene regulatory mechanisms in an indirect manner as these two molecules may act as host for other sncRNAs. In my thesis, I captured some of the piRNAs and miRNAs embedded within snoRNAs and tRNAs. I also observed that these RNAs exhibit same direction of expression in the tumor tissues as that of the snoRNAs. The corresponding gene targets for these piRNAs and miRNAs were also identified but needs to be functionally ascertained in

luciferase expression constructs carrying 3' UTRs of target genes and sncRNA of interest. Previous findings from literature has also suggested that dysregulation of snoRNAs may lead to phenotypic changes. We therefore need to identify key snoRNA molecules and understand this behavior in breast cancer cell lines. We also need to identify if the phenotypic change is because of the presence of embedded regulatory sncRNAs or due to snoRNA/tRNA per se. However, transfection of tRNAs into cell lines and understanding their effect on phenotypic changes is challenging and has not been attempted so far because of the structural intricacies that these molecules possess. Stable transfection constructs carrying specific tRNAs may potentially overcome some of the barriers in addressing gene-phenotype relationships.

d. Gene targets identified for piRNAs and miRNAs were *in silico* based which were checked for their expression in tissue datasets. However, to confirm their direct regulation, the first step is to perform a luciferase reporter assay and establish their direct regulation. Subsequently, one needs to check the protein expression levels using western blots. Further understanding on the impact of gene regulation to cell phenotype has to be carried out using appropriate cell based assays.

8.4 Conclusions

The challenge in breast cancer management lies in accurately identifying patients who are likely to undergo recurrence/death and in knowledge transfer from the biomarker studies to the clinical setting to improve responses to therapies or survival. Several attempts have been made for this purpose using gene expression but have often met with limited success. Therefore, the search for biomarkers continues and this thesis was aimed

at identifying other biomolecules that could serve as biomarkers for breast cancer. I focused on small non-coding RNAs and identified many RNAs that showed promise for breast cancer prognosis. Two novel miRNAs were identified, apart from validating the prognostic relevance of ten other miRNAs. Our understanding on the clinical relevance and the functional roles of piRNAs and PIWI genes are, at best, rudimentary. This study has identified piRNAs and PIWI genes that have not been previously reported for breast cancer. The role of tRNAs as prognostic markers for any cancer type is unheard of, and this study, I believe, will open up more research in this domain. The association of snoRNAs with breast cancer is also new. This work on genome-wide profiling of sncRNAs using modern sequencing platforms significantly augments the limited previous literature, and the data provided in this study therefore extends the comprehensive search for breast cancer biomarkers. The key to advance promising biomarkers to clinical trials is by making profiling and quantifying of these sncRNAs easy, establishing gold standard approaches to profile and analyze these molecules, gaining functional insights into the identified biomarkers and last but not the least, is through independent confirmation of findings, independent of confounding variables (histological and molecular heterogeneity, stage, grade and treatments).

8.5 References

1. Kovalchuk O, Filkowski J, Meservy J, Ilnytskyy Y, Tryndyak VP, Chekhun VF, et al. Involvement of microRNA-451 in resistance of the MCF-7 breast cancer cells to chemotherapeutic drug doxorubicin. Mol Cancer Ther 2008;7:2152-2159.

2. He L, Hannon GJ. MicroRNAs: Small RNAs with a big role in gene regulation. Nature Reviews Genetics 2004;5:522-531.

3. Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. ONCOTARGET 2014;5:9901-10.

4. Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, et al. Identification of novel piRNAs in bladder cancer. Cancer Lett 2015;356:561-567.

5. Martens-Uzunova E, Olvedy M, Jenster G. Beyond microRNA - novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett 2013;340:201-211.

6. Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, et al. Article: A human snoRNA with MicroRNA-like functions. Mol Cell 2008;32:519-528.

7. Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res 2011;39:675-86.

8. Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, et al. A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. Nucleic Acids Res 2015;43:10474-10491.

9. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 2013;110:1404-9.

10. Okugawa Y, Toiyama Y, Toden S, Mitoma H, Nagasaka T, Tanaka K, et al. Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. Gut 2015.

11. Wei R, Huang G, Zhang M, Li B, Zhang H, Shi M, et al. Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. Clin Cancer Res 2013;19:4780-4791.

12. Schummer M, Green A, Beatty JD, Karlan BY, Karlan S, Gross J, et al. Comparison of breast cancer to healthy control tissue discovers novel markers with potential for prognosis and early detection. PLoS ONE 2010;5:1-9.

13. Sanz-Pamplona R, Berenguer A, Cordero D, Mollev? DG, Crous-Bou M, Sole X, et al. Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. Molecular Cancer 2014;13:46.

14. Braakhuis BJM, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of slaughter's concept of field cancerization: Evidence and clinical implications. Cancer Res 2003;63:1727-1730.

15. Hawthorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. Genomics 2014;103:211-221.

16. Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC Cancer 2005;5:45-11.

17. Troester MA, Lee MH, Carter M, Fan C, Cowan DW, Perez ER, et al. Activation of host wound responses in breast cancer microenvironment. Clin Cancer Res 2009;15:7020-7028.

18. Radovich M, Clare SE, Atale R, Pardo I, Hancock BA, Solzak JP, et al. Characterizing the heterogeneity of triple-negative breast cancers using microdissected normal ductal epithelium and RNA-sequencing. Breast Cancer Res Treat 2014;143:57-68.

19. Hannafon BN, Sebastiani P, de lM, Lu J, Rosenberg CL. Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer. Breast Cancer Res 2011;13:R24-R24.

20. Wagner E, Liu Y, Schneider B, Storniolo AM, Han J, He C. Abstract 3268: Differences in expression of lncRNAs in breast tumor, adjacent normal-appearing breast tissue, and normal breast tissue from healthy donors. Cancer Research 2014;74:3268-3268.

21. Zhang X, Chen J, Radcliffe T, Lebrun DP, Tron VA, Feilotter H. An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. J Mol Diagn 2008;10:513-519.

22. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol 2007;7:36-36.

23. Ludyga N, Grünwald B, Azimzadeh O, Englert S, Höfler H, Tapio S, et al. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. VIRCHOWS ARCHIV 2012;460:131-140.

24. Kong H, Zhu M, Cui F, Wang S, Gao X, Lu S, et al. Quantitative assessment of short amplicons in FFPE-derived long-chain RNA. Sci Rep 2014;4:7246-7246.

25. Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. Microarray-based, high-throughput gene expression profiling of microRNAs. Nature Methods 2004;1:155-161.

26. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. RNA 2007;13:1668-1674.

27. Bovell L, Shanmugam C, Katkoori VR, Zhang B, Vogtmann E, Grizzle WE, et al. microRNAs are stable in formalin-fixed paraffin-embedded archival tissue specimens of colorectal cancers stored for up to 28 years. Frontiers in bioscience (Elite edition) 2012;4:1937-1940.

28. Peiró-Chova L, Peña-Chilet M, López-Guerrero JA, García-Giménez JL, Alonso-Yuste E, Burgues O, et al. High stability of microRNAs in tissue samples of compromised quality. Virchows Archiv 2013;463:765-774.

29. Liu A, Tetzlaff MT, Vanbelle P, Elder D, Feldman M, Tobias JW, et al. MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues. Int J Clin Exp Pathol 2009;2:519-527.

30. Tam S, de Borja R, Tsao M, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest 2014;94:350-358.

31. Liu J, Jennings SF, Tong W, Hong H. Next generation sequencing for profiling expression of miRNAs: Technical progress and applications in drug development. Journal of Biomedical Science and Engineering (JBiSE) 2011;4:666-676.

32. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: Approaches and considerations. Nature Reviews Genetics 2012;13:358-369.

33. Meng W, McElroy JP, Volinia S, Palatini J, Warner S, Ayers LW, et al. Comparison of MicroRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. PLoS ONE 2013;8:1-9.

34. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010;11:733-9.

35. Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG. Quantification of microRNA expression with next-generation sequencing. Curr Protoc Mol Biol 2013;Chapter 4:4.17-4.17.

36. Nygaard S, Jacobsen A, Lindow M, Eriksen J, Balslev E, Flyger H, et al. Identification and analysis of miRNAs in human breast cancer and teratoma samples using deep sequencing. BMC Medical Genomics 2009;2:35.

37. Schee K, Lorenz S, Worren MM, Günther C, Holden M, Hovig E, et al. Deep sequencing the MicroRNA transcriptome in colorectal cancer. PLoS ONE 2013;8:1-11.

38. Chang H, Li S, Ho M, Pan H, Ger L, Hu L, et al. Comprehensive analysis of microRNAs in breast cancer. BMC Genomics 2012;13 Suppl 7:S18-S18.

39. Lopez JP, Diallo A, Cruceanu C, Fiori LM, Laboissiere S, Guillet I, et al. Biomarker discovery: Quantification of microRNAs and other small non-coding RNAs using next generation sequencing. BMC Med Genomics 2015;8:35-35.

40. Xiong J, Bing Z, Su Y, Deng D, Peng X. An integrated mRNA and microRNA expression signature for glioblastoma multiforme prognosis. PLoS ONE 2014;9:1-8.

41. Harrell FE, J., Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-387.

42. Farazi TA, Horlings HM, Ten Hoeve J,J., Mihailovic A, Halfwerk H, Morozov P, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res 2011;71:4443-4453.

43. Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc Natl Acad Sci U S A 2015;112:E1106-E1115

44. Tatarano S, Chiyomaru T, Kawakami K, Enokida H, Yoshino H, Hidaka H, et al. Novel oncogenic function of mesoderm development candidate 1 and its regulation by MiR-574-3p in bladder cancer cell lines. Int J Oncol 2012;40:951-959.

45. Ujihira T, Ikeda K, Suzuki T, Yamaga R, Sato W, Horie-Inoue K, et al. MicroRNA-574-3p, identified by microRNA library-based functional screening, modulates tamoxifen response in breast cancer. Sci Rep 2015;5:7641-7641.

Bibliography

Abe O, A., Abe R, A., Enomoto K, A., Kikuchi K, A., Koyama H, A., Masuda H, A., et al. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. Lancet 2005:1687-1717.

American Cancer Society. Breast Cancer Facts & Figures 2015-2016. Atlanta: American Cancer Society, Inc. 2015; Available from: http://www.cancer.org/research/cancerfactsstatistics/breast-cancer-facts-figures [Cited 2016 Jun].

Aparicio S, Mardis E. Tumor heterogeneity: Next-generation sequencing enhances the view from the pathologist's microscope. Genome Biol 2014;15:463-463.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 2006;442:203-207.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 2008;31:785-799.

Arranz EE, Vara JÁF, Gámez-Pozo A, Zamora P. Gene signatures in breast cancer: Current and future uses. Translational Oncology 2012;5:398-403.

Assumpcao CB, Calcagno DQ, Araujo T, dos Santos S, dos Santos A, Riggins GJ, et al. The role of piRNA and its potential clinical implications in cancer. Epigenomics 2015;7:975-984.

Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. Women's Oncology Review 2004;4:211-212.

Bachellerie JP, Cavaille J, Huttenhofer A. The expanding snoRNA world. Biochimie 2002;84:775-790.

Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. Science 1986;234:179-186.

Bahn JH, Zhang Q, Li F, Chan TM, Lin X, Kim Y, et al. The landscape of MicroRNA, piwi-interacting RNA, and circular RNA in human saliva. Clin Chem 2015;61:221-230.

Baker M. MicroRNA profiling: Separating signal from noise: Various platforms for measuring microRNAs can provide different answers. Nature Methods 2010;7:687-692.

Bartel DP. Review: MicroRNAs. genomics, biogenesis, mechanism, and function. Cell 2004;116:281-297.

Bastani M, Vos L, Asgarian N, Deschenes J, Graham K, Mackey J, et al. A machine learned classifier that uses gene expression data to accurately predict estrogen receptor status. PLoS one 2013;8:1-9.

Berindan-Neagoe I, Monroig PD, Pasculli B, Calin GA. MicroRNAome genome: A treasure for cancer diagnosis and therapy. Ca-A cancer journal for clinicians 2014;64:311-336.

Berindan-Neagoe I, Monroig PD, Pasculli B, Calin GA. MicroRNAome genome: A treasure for cancer diagnosis and therapy. CA-A Cancer journal for clinicians 2014;64:311-36.

Bermudez-Santana C, Attolini C, Kirsten T, Engelhardt J, Prohaska SJ, Steigele S, et al. Genomic organization of eukaryotic tRNAs. BMC Genomics 2010;11:270.

Bernstein L, Lacey,James V.,,Jr. Receptors, associations, and risk factor differences by breast cancer subtypes: Positive or negative? J Natl Cancer Inst 2011;103:451-453.

Bertoli G, Cava C, Castiglioni I. MicroRNAs: New biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. Theranostics 2015;5:1122-1143.

Bertucci F, Birnbaum D. Reasons for breast cancer heterogeneity. J Biol 2008;7:6-6.

Bjornsti MA, Houghton PJ. Lost in translation: Dysregulation of cap-dependent translation and cancer. Cancer cell 2004;5:519-523.

Bogaerts J, Cardoso F, Buyse M, Braga S, Loi S, Harrison JA, et al. Gene signature evaluation as a prognostic tool: Challenges in the design of the MINDACT trial. Nature clin pract oncol 2006;3:540-551.

Bogdanova N, Helbig S, Dörk T. Hereditary breast cancer: Ever more pieces to the polygenic puzzle. Hereditary Cancer in Clinical Practice 2013;11:1-16.

Bombardieri E, Gianni L, Bonadonna G, Fabbri A, Carcangiu ML, Carbone A. Histological classification of breast cancer. Breast Cancer - Nuclear Medicine in Diagnosis and Therapeutic Options 2008; p. 3.

Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nature Structural & Molecular Biology 2006;13:1097-1101.

Bovell L, Shanmugam C, Katkoori VR, Zhang B, Vogtmann E, Grizzle WE, et al. microRNAs are stable in formalin-fixed paraffin-embedded archival tissue specimens of colorectal cancers stored for up to 28 years. Frontiers in bioscience (Elite edition) 2012;4:1937-1940.

Braakhuis BJM, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of slaughter's concept of field cancerization: Evidence and clinical implications. Cancer Res 2003;63:1727-1730.

Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res 2011;39:675-686.

Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res 2011;39:675-86.

Breast cancer facts & figures. [electronic resource]. : Atlanta, Ga. : American Cancer Society [Cited 2016 June].

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in drosophila. Cell 2007;128:1089-1103.

Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, et al. Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev 2007;21:2300-2311.

Busch J, Ralla B, Jung M, Wotschofsky Z, Trujillo-Arribas E, Schwabe P, et al. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. Journal of Experimental & Clinical Cancer Research (17569966) 2015;34:1-11.

Busk PK. A tool for design of primers for microRNA-specific quantitative RT-qPCR. BMC Bioinformatics 2014;15:2-22.

Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183-1192.

Cai J, Guan H, Fang L, Yang Y, Zhu X, Yuan J, et al. MicroRNA-374a activates wnt/ß-catenin signaling to promote breast cancer metastasis. J Clin Invest 2013;123:566-579.

Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 2002;99:15524-15529.

Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2015. Toronto, ON: Canadian Cancer Society; 2015. 2015; Available from: http://www.cancer.ca/en/cancer-information/cancer-101/canadian-cancer-statistics-publication/?region=on [Cited 2016 Jun].

Cao JN. The functional role of long non-coding RNAs and epigenetics. Biol Proc Online 2014;16:11.

Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. The argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev 2002;16:2733-2742.

Catalanotto C, Azzalin G, Macino G, Cogoni C. Involvement of small RNAs and role of the qde genes in the gene silencing pathway in neurospora. Genes Dev 2002;16:790-795.

Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res 2005;65:6029-6033.

Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 2013;19:4477-4487.

Chan PP, Lowe TM. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 2009;37:D93-D97.

Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC Cancer 2005;5:45-11.

Chang H, Li S, Ho M, Pan H, Ger L, Hu L, et al. Comprehensive analysis of microRNAs in breast cancer. BMC Genomics 2012;13 Suppl 7:S18-S18.

Chang HY, Dimitry SAN, Sneddon JB, Hastie T, Tibshirani R, Sørlie T, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738-3743.

Chen C, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science 2004;303:83-86.

Chen L, Han L, Wei J, Zhang K, Shi Z, Duan R, et al. SNORD76, a box C/D snoRNA, acts as a tumor suppressor in glioblastoma. Sci Rep 2015;5:8588-8588.

Chen W, Böcker W, Brosius J, Tiedge H. Expression of neural BC200 RNA in human tumours. J Pathol 1997;183:345-351.

Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008;18:997-1006.

Cheng J, Guo J, Xiao B, Miao Y, Jiang Z, Zhou H, et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. Clinica Chimica Acta 2011;412:1621-1625.

Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, et al. Identification of novel piRNAs in bladder cancer. Cancer Lett 2015;356:561-567.

Chu L, Su MY, Maggi L,Jr, Lu L, Mullins C, Crosby S, et al. Multiple myeloma-associated chromosomal translocation activates orphan snoRNA ACA11 to suppress oxidative stress. J Clin Invest 2012;122:2793-2806.

Ciechanover A, Wolin SL, Steitz JA, Lodish HF. Transfer RNA is an essential component of the ubiquitin- and ATP-dependent proteolytic system. Proc Natl Acad Sci U S A 1985:1341.

CIendening JW, Pandyra A, Boutros PC, Ghamrasni SE, Khosravi F, Trentin GA, et al. Dysregulation of the mevalonate pathway promotes transformation. Proc Natl Acad Sci U S A 2010;107:15051-15056.

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A 2005;102:13944-13949.

Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: International survey. BMJ 2000;320:1240-1243.

Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev 1998;12:3715-3727.

Creemers EE, Tijsen AJ, Pinto YM. Circulating MicroRNAs novel biomarkers and extracellular communicators in cardiovascular disease? Circ Res 2012;110:483-495.

Crick FH. Codon--anticodon pairing: The wobble hypothesis. J Mol Biol 1966;19:548-555.

Cui L, Lou Y, Zhang X, Zhou H, Deng H, Song H, et al. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. Clin Biochem 2011;44:1050-1057.

Cuk K, Zucknick M, Heil Jö, Madhavan D, Schott S, Turchinovich A, et al. Circulating microRNAs in plasma as early detection markers for breast cancer. Int J Cancer 2013;132:1602-1612.

Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the genomic health recurrence score in early breast cancer. J Clin Oncol 2011;29:4273-4278.

Dabbs DJ, Klein ME, Mohsin SK, Tubbs RR, Shuai Y, Bhargava R. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the oncotype DX test: An independent quality assurance study. J Clin Oncol 2011;29:4279-4285.

Damaraju S, Driga A, Cook L, Calder K, Graham K, Dabbs K, et al. 3.02: Considerations on dealing with tissues and cell samples (include tissue banking). Comprehensive Sampling and Sample Preparation 2012:21-31.

Davis LM, Harris C, Tang L, Doherty P, Hraber P, Sakai Y, et al. Amplification patterns of three genomic regions predict distant recurrence in breast carcinoma. Journal of molecular diagnostics 2007;9:327-336.

de Boer M, van Dijck JA, Bult P, Borm GF, Tjan-Heijnen VC. Breast cancer prognosis and occult lymph node metastases, isolated tumor cells, and micrometastases. J Natl Cancer Inst 2010;102:410-425.

De Cecco L, Dugo M, Canevari S, Daidone MG, Callari M. Measuring microRNA expression levels in oncology: From samples to data analysis. Crit Rev Oncog 2013;18:273-287.

de Haro M, Al-Ramahi I, Jones KR, Holth JK, Timchenko LT, Botas J. Smaug/SAMD4A restores translational activity of CUGBP1 and suppresses CUG-induced myopathy. PLoS Genetics 2013;9:1-10.

Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nature Reviews Cancer 2005:275.

Deng W, Lin HF. Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Developmental cell 2002;2:819-830.

Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep sequencing of serum small RNAs identifies patterns of 5' tRNA half and YRNA fragment expression associated with breast cancer. Biomark Cancer 2014;6:37-47.

Dieci G, Preti M, Montanini B. Minireview: Eukaryotic snoRNAs: A paradigm for gene expression flexibility. Genomics 2009;94:83-88.

Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. PLoS Genetics 2006;3:2107-2115.

Dobbin KK, Simon RM. Sample size planning for developing classifiers using high-dimensional DNA microarray data. Biostatistics 2007;8:101-117.

Dobbin KK, Zhao Y, Simon RM. How large a training set is needed to develop a classifier for microarray data? Clin Cancer Res 2008;14:108-114.

Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. Genes Dev 2004;18:504-511.

Dolfi SC, Chan LL, Qiu J, Tedeschi PM, Bertino JR, Hirshfield KM, et al. The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. Cancer & Metabolism 2013;1:20.

Dong J, Qiu H, Garcia-Barrio M, Anderson J, Hinnebusch AG. Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. Mol Cell 2000;6:269-279.

Dong X, Guo P, Boyd J, Sun X, Li Q, Zhou W, et al. Implication of snoRNA U50 in human breast cancer. Journal of Genetics and Genomics 2009;36:447-454.

Dragon F, Filipowicz W. SnoRNAs: Biogenesis, structure and function. Encyclopedia of Life Sciences 2005:317.

Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG. Quantification of microRNA expression with next-generation sequencing. Curr Protoc Mol Biol 2013;Chapter 4:4.17-4.17.

Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, et al. Article: A human snoRNA with MicroRNA-like functions. Mol Cell 2008;32:519-528.

Engstrøm,M.J., Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. Breast Cancer Res Treat 2013;140:463-473.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in drosophila. Genome Biol 2003;5:R1-R1.

Eroles P, Bosch A, Alejandro Pérez-Fidalgo J, Lluch A. Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. Cancer Treat Rev 2012;38:698-707.

Esposito T, Magliocca S, Formicola D, Gianfrancesco F. piR_015520 belongs to piwi-associated RNAs regulates expression of the human melatonin receptor 1A gene. PLOS ONE 2011;6:e22727.

Falkenberg N, Anastasov N, Rappl K, Braselmann H, Auer G, Walch A, et al. MiR-221/-222 differentiate prognostic groups in advanced breast cancers and influence cell invasion. Br J Cancer 2013;109:2714-2723.

Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. Lancet 2001;358 North American Edition:1389-1399.

Farazi TA, Horlings HM, Ten Hoeve J,J., Mihailovic A, Halfwerk H, Morozov P, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res 2011;71:4443-4453.

Fazi F, Nervi C. MicroRNA: Basic mechanisms and transcriptional regulatory networks for cell fate determination. Cardiovasc Res 2008;79:553-561.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11

[Internet]. Lyon, France: International Agency for Research on Cancer. 2013; Available from: http://globocan.iarc.fr [Cited 2016 Jun].

Filipowicz W, Pogacic V. Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol 2002;14:319-327.

Gall JG. The centennial of the cajal body. Nature Reviews Molecular Cell Biology 2003;4:975-980.

Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, et al. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. Int J Cancer 2015;136:E623-E629.

Garber K. Gene expression tests foretell breast cancer's future. Science 2004:1754-1755.

Garcia AI, Buisson M, Bertrand P, Rimokh R, Rouleau E, Lopez BS, et al. Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. EMBO Mol Med 2011;3:279-290.

García-Closas M, Brinton LA, Lissowska J, Chatterjee N, Peplonska B, Anderson WF, et al. Established breast cancer risk factors by clinically important tumour characteristics. Br J Cancer 2006;95:123-129.

Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-1184.

Germain DR, Graham K, Glubrecht DD, Hugh JC, Mackey JR, Godbout R. DEAD box 1: A novel and independent prognostic marker for early recurrence in breast cancer. Breast Cancer Res Treat 2011;127:53-63.

Girard A, Hannon GJ. Conserved themes in small-RNA-mediated transposon control. Trends Cell Biol 2008;18:136-148.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian piwi proteins. Nature 2006;442:199-202.

GLOBOCAN 2012 v 1.0, cancer incidence and mortality worldwide: IARC CancerBase no. 11. Journal of Endourology 2015;29:495-497.

Goetz MP, Suman VJ, Ingle JN, Nibbe AM, Visscher DW, Reynolds CA, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. Clinical cancer research 2006;12:2080-2087.

Gogol-Döring A, Chen W. An overview of the analysis of next generation sequencing data. Methods Mol Biol 2012;802:249-257.

Gökmen-Polar Y, Badve S. Breast cancer prognostic markers: Where are we now? MLO: Medical Laboratory Observer 2012;44:22-25.

Gonzalez-Angulo A, Morales-Vasquez F, Hortobagyi GN. Overview of resistance to systemic therapy in patients with breast cancer. Adv Exp Med Biol 2007;608:1-22.

Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. Cell 2015;161:790-802.

Gospodarowicz MK. Prognostic factors in cancer. : New York : Wiley-Liss, c2001; 2nd ed; 2001.

Grada A, Weinbrecht K. Next-generation sequencing: Methodology and application. J Invest Dermatol 2013;133:E1-E4.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006;34:D140-D144.

Griffiths-Jones S. The microRNA registry. Nucleic Acids Res 2004;32:D109-D111.

Grivna ST, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20:1709-14.

Grivna ST, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20:1709-1714.

Grochola LF, Greither T, Taubert H, Moller P, Knippschild U, Udelnow A, et al. The stem cell-associated hiwi gene in human adenocarcinoma of the pancreas: Expression and risk of tumour-related death. Br J Cancer 2008;99:1083-1088.

Guo Y, Bosompem A, Mohan S, Erdogan B, Ye F, Vickers KC, et al. Transfer RNA detection by small RNA deep sequencing and disease association with myelodysplastic syndromes. BMC Genomics 2015;16:727.

Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013;4:627-635.

Hall JS, Taylor J, Valentine HR, Irlam JJ, Eustace A, Hoskin PJ, et al. Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. Br J Cancer 2012;107:684-694.

Hannafon BN, Sebastiani P, de lM, Lu J, Rosenberg CL. Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer. Breast Cancer Res 2011;13:R24-R24.

Haque R, Ahmed SA, Inzhakova G, Shi J, Avila C, Polikoff J, et al. Impact of breast cancer subtypes and treatment on survival: An analysis spanning two decades. Cancer Epidemiology, Biomarkers & Prevention 2012;21:1848-1855.

Harrell FE, J., Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-387.

Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-9910.

Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, Nassa G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. Oncotarget 2014;5:9901-10.

Hausser J, Zavolan M. Identification and consequences of miRNA-target interactions--beyond repression of gene expression. Nature Reviews Genetics 2014;15:599-612.

Hawthorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. Genomics 2014;103:211-221.

He G, Chen L, Ye Y, Xiao Y, Hua KD, Jarjoura D, et al. Piwil2 expressed in various stages of cervical neoplasia is a potential complementary marker for p16(INK4a). American journal of translational research 2010;2:156-169.

He L, Hannon GJ. MicroRNAs: Small RNAs with a big role in gene regulation. Nature Reviews Genetics 2004;5:522-531.

Hoagland MB, keller EB, zamecnik PC. Enzymatic carboxyl activation of amino acids. J Biol Chem 1956;218:345-358.

Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC. A soluble ribonucleic acid intermediate in protein synthesis. J Biol Chem 1958;231:241-257.

Hong L, Yang J, Han Y, Lu Q, Cao J, Syed L. High expression of miR-210 predicts poor survival in patients with breast cancer: A meta-analysis. Gene 2012;507:135-138.

Hopper AK, Phizicky EM. tRNA transfers to the limelight. Genes Dev 2003;17:162-180.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van dE, et al. Article: A role for piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell 2007;129:69-82.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009;4:44-57.

Huang G, Hu H, Xue X, Shen S, Gao E, Guo G, et al. Altered expression of piRNAs and their relation with clinicopathologic features of breast cancer. Clinical & translational oncology 2013;15:563-568.

Hummel R, Sie C, Watson DI, Wang T, Ansar A, Michael MZ, et al. MicroRNA signatures in chemotherapy resistant esophageal cancer cell lines. World J Gastroenterol 2014;20:14904-14912.

Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief Funct Genomic Proteomic 2009;8:174-183.

Hutchinson L. Breast cancer: Challenges, controversies, breakthroughs. Nat Rev Clin Oncol 2010;7:669-670.

Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065-70.

Italiano A. Prognostic or predictive? it's time to get back to definitions! J Clin Oncol 2011;29:4718; author reply 4718-4719.

Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: Its biogenesis and functions. Annu Rev Biochem 2015;84:405-433.

Janssen HLA, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K, et al. Treatment of HCV infection by targeting microRNA. N Engl J Med 2013;368:1685-1694.

Jian-Fu Chen, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 2006;38:228-233.

Jiang X, Du L, Wang L, Li J, Liu Y, Zheng G, et al. Serum microRNA expression signatures identified from genome-wide microRNA profiling serve as novel noninvasive biomarkers for diagnosis and recurrence of bladder cancer. Int J Cancer 2015;136:854-862.

Keam SP, Hutvagner G. tRNA-derived fragments (tRFs): Emerging new roles for an ancient RNA in the regulation of gene expression. Life (2075-1729) 2015;5:1638-1651.

Keam SP, Young PE, McCorkindale AL, Dang THY, Clancy JL, Humphreys DT, et al. The human piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. Nucleic Acids Res 2014;42:8984-8995.

Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl genomes 2016: More genomes, more complexity. Nucleic Acids Res 2016;44:D574-D580.

Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: Reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94:606-616.

Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet oncol 2001;2:133-140.

Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and rniRNAs exhibit strand bias. Cell 2003;115:209-216.

Kim B, Hong SW, Kim A, Choi SH, Yoon SO. Prognostic implications for high expression of oncogenic microRNAs in advanced gastric carcinoma. J Surg Oncol 2013;107:505-510.

Kim J, Inoue K, Ishii J, Vanti WB, Voronov SV, Murchison E, et al. A MicroRNA feedback circuit in midbrain dopamine neurons. Science 2007;317:1220-1224.

Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nature Reviews Molecular Cell Biology 2009:126.

Kim VN. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. Genes Dev 2006;20:1993-1997.

Kishore S, Stamm S. Regulation of alternative splicing by snoRNAs. Cold Spring Harb Symp Quant Biol 2006;71:329-334.

Kleivi Sahlberg K, Bottai G, Naume B, Burwinkel B, Calin GA, Børresen-Dale A, et al. A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. Clin Cancer Res 2015;21:1207-1214.

Ko A, Dollinger M, Rosenbaum EH. Everyone's guide to cancer therapy : How cancer is diagnosed, treated, and managed day to day. : Kansas City : Andrews McMeel Pub., c2008; Rev. 5th ed; 2008.

Kodahl AR, Lyng MB, Binder H, Cold S, Gravgaard K, Knoop AS, et al. Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: A case control study. Molecular Oncology 2014;8:874-883.

Kolbert CP, Feddersen RM, Rakhshan F, Grill DE, Simon G, Middha S, et al. Multi-platform analysis of MicroRNA expression measurements in RNA from fresh frozen and FFPE tissues. PLoS ONE 2013;8:1-9.

Kong H, Zhu M, Cui F, Wang S, Gao X, Lu S, et al. Quantitative assessment of short amplicons in FFPE-derived long-chain RNA. Sci Rep 2014;4:7246-7246.

Kovalchuk O, Filkowski J, Meservy J, Ilnytskyy Y, Tryndyak VP, Chekhun VF, et al. Involvement of microRNA-451 in resistance of the MCF-7 breast cancer cells to chemotherapeutic drug doxorubicin. Mol Cancer Ther 2008;7:2152-2159.

Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37:495-500.

Krishnan P, Ghosh S, Graham K, Mackey J, Kovalchuk O, Damaraju S. Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer. Oncotarget 2016.

Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics 2015;16:735-735.

Kuramochi-Miyagawa S, Kimura T, Yomogida K, Kuroiwa A, Tadokoro Y, Fujita Y, et al. Two mouse piwi-related genes: Miwi and mili. Mech Dev 2001;108:121-133.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001;294:853-858.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25-R25.

Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in caenorhabditis elegans. Science 2001;294:858-862.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science 2006;313:363-367.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science 2006:313:363-367.

Le Thomas A, Tóth KF, Aravin AA. To be or not to be a piRNA: Genomic origin and processing of piRNAs. Genome Biol 2014;15:204.

Lee JH, Jung C, Javadian-Elyaderani P, Schweyer S, Schütte D, Shoukier M, et al. Pathways of proliferation and antiapoptosis driven in breast cancer stem cells by stem cell protein piwil2. Cancer Res 2010;70:4569-4579.

Lee JH, Schütte D, Wulf G, Füzesi L, Radzun H, Schweyer S, et al. Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/bcl-XL pathway. Hum Mol Genet 2006;15:201-211.

Lee RC, Ambros V. An extensive class of small RNAs in caenorhabditis elegans. Science 2001:862-864.

Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 2004;116:843-854.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III drosha initiates microRNA processing. Nature 2003;425:415-419.

Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: Stepwise processing and subcellular localization. EMBO J 2002;21:4663-4670.

Lee YS, Lim YS, Lee J, Wang S, Park H, Kim SY, et al. Differential expression levels of plasma-derived miR-146b and miR-155 in papillary thyroid cancer. Oral Oncol 2015;51:77-83.

Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 2009;23:2639-2649.

Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 2009;23:2639-49.

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010;11:733-739.

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010;11:733-9.

Li J, Kong F, Wu K, Song K, He J, Sun W. miR-193b directly targets STMN1 and uPA genes and suppresses tumor growth and metastasis in pancreatic cancer. Molecular Medicine Reports 2014;10:2613-2620.

Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol 2007;7:36-36.

Li J, Wang Y, Yu W, Chen J, Luo J. Expression of serum miR-221 in human hepatocellular carcinoma and its prognostic significance. Biochem Biophys Res Commun 2011;406:70-73.

Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997:1943-1947.

Li X-, Yan P-, Shao Z-. Downregulation of miR-193b contributes to enhance urokinase-type plasminogen activator (uPA) expression and tumor progression and invasion in human breast cancer. Oncogene 2009;28:3937-3948.

Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-microRNA signature. Gut 2010;59:579-585.

Li XJ, Ren ZJ, Tang JH. MicroRNA-34a: A potential therapeutic target in human cancer. Cell death & disease 2014;5:e1327.

Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Molecular Cancer 2010;9:198-207.

Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkersfor non-small-cell lung cancer. Molecular Cancer 2010;9:198-207.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer: Analyses of cohorts of twins from sweden, denmark, and finland. N Engl J Med 2000;343:78-85.

Liu A, Tetzlaff MT, Vanbelle P, Elder D, Feldman M, Tobias JW, et al. MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues. Int J Clin Exp Pathol 2009;2:519-527.

Liu D, Zhao H, Zhao S, Wang X. MicroRNA expression profiles of peripheral blood mononuclear cells in patients with systemic lupus erythematosus. Acta Histochem 2014; 116:891-897.

Liu J, Jennings SF, Tong W, Hong H. Next generation sequencing for profiling expression of miRNAs: Technical progress and applications in drug development. Journal of Biomedical Science and Engineering (JBiSE) 2011;4:666-676.

Liu JJ, Shen RL, Chen L, Ye Y, He G, Hua KD, et al. Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. International journal of clinical and experimental pathology 2010;3:328-337.

Liu N, Cui R, Sun Y, Guo R, Mao Y, Tang L, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. International Journal of Cancer 2014;134:1359-68.

Liu N, Cui R, Sun Y, Guo R, Mao Y, Tang L, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. International Journal of Cancer 2014;134:1359-1368.

Liu N, Cui R, Sun Y, Guo R, Mao Y, Tang L, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. International Journal of Cancer 2014:1359.

Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217-226.

Liu X, He M, Hou Y, Liang B, Zhao L, Ma S, et al. Expression profiles of microRNAs and their target genes in papillary thyroid carcinoma. Oncol Rep 2013;29:1415-1420.

Liu XY, Sun Y, Guo JP, Ma H, Li JY, Dong B, et al. Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. International journal of cancer 2006;118:1922-1929.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 2001;25:402-408.

Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc Natl Acad Sci U S A 2015;112:E1106-E1115

Lønning,P.E. Breast cancer prognostication and prediction: Are we making progress? Ann Oncol 2007;18 Suppl 8:viii3-viii7.

Lopes-Ramos C, Habr-Gama A, Quevedo BD, Felcio NM, Bettoni F, Koyama FC, et al. Overexpression of miR-21-5p as a predictive marker for complete tumor regression to neoadjuvant chemoradiotherapy in rectal cancer patients. Bmc medical genomics 2014;7:68.

Lopez JP, Diallo A, Cruceanu C, Fiori LM, Laboissiere S, Guillet I, et al. Biomarker discovery: Quantification of microRNAs and other small non-coding RNAs using next generation sequencing. BMC Med Genomics 2015;8:35-35.

Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955-64.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834-838.

Ludyga N, Grünwald B, Azimzadeh O, Englert S, Höfler H, Tapio S, et al. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. VIRCHOWS ARCHIV 2012;460:131-140.

Luo J, Chen M, Huang H, Yuan T, Zhang M, Zhang K, et al. Circulating microRNA-122a as a diagnostic marker for hepatocellular carcinoma. OncoTargets & Therapy 2013;6:577-583.

Luteijn MJ, Ketting RF. PIWI-interacting RNAs: From generation to transgenerational epigenetics. Nature Reviews Genetics 2013;14:523-534.

Malkin D, Li FP, Strong LC, Fraumeni, Joseph F., Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 1990:1233-1238.

Marshall L, White RJ. Non-coding RNA production by RNA polymerase III is implicated in cancer. Nature Reviews Cancer 2008;8:911-914.

Marshall L, White RJ. Non-coding RNA production by RNA polymerase III is implicated in cancer. Nature Reviews Cancer 2008:911.

Martens-Uzunova E, Olvedy M, Jenster G. Beyond microRNA - novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett 2013;340:201-211.

Martinez VD, Enfield KSS, Rowbotham DA, Lam WL. An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence. Gastric Cancer 2016;19:660-665.

Martinez VD, Enfield KSS, Rowbotham DA, Lam WL. An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence. Gastric Cancer 2015;19:660-665.

Martinez VD, Vucic EA, Thu KL, Hubaux R, Enfield KSS, Pikor LA, et al. Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. Sci Rep 2015;5:10423-10423.

Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 2013;110:1404-1409.

Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 2013;110:1404-9.

McPherson K, Steel CM, Dixon JM. ABC of breast diseases: Breast cancer -- epidemiology, risk factors, and genetics. BMJ 2000;321:624-628.

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Guidelines: REporting recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer 2005;41:1690-1696.

Mei Y, Liao J, Shen J, Yu L, Liu B, Liu L, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. Oncogene 2012;31:2794-2804.

Mei YD, Yong J, Liu HT, Shi YG, Meinkoth J, Dreyfuss G, et al. tRNA binds to cytochrome c and inhibits caspase activation. Mol Cell 2010;37:668-678.

Mei YD, Yong J, Liu HT, Shi YG, Meinkoth J, Dreyfuss G, et al. tRNA binds to cytochrome c and inhibits caspase activation. Mol Cell 2010;37:668-78.

Mei YP, Clark D, Mao L. Novel dimensions of piRNAs in cancer. Cancer Lett 2013;336:46-52.

Meng W, McElroy JP, Volinia S, Palatini J, Warner S, Ayers LW, et al. Comparison of MicroRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. PLoS ONE 2013;8:e64393.

Meng W, McElroy JP, Volinia S, Palatini J, Warner S, Ayers LW, et al. Comparison of MicroRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. PLoS ONE 2013;8:1-9.

Merlos-Suarez A, Barriga FM, Jung P, Iglesias M, Cespedes MV, Rossell D, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell stem cell 2011;8:511-524.

Michel CI, Holley CL, Scruggs BS, Sidhu R, Brookheart RT, Listenberger LL, et al. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. Cell metabolism 2011;14:33-44.

Miller BH, Zeier Z, Xi L, Lanz TA, Deng S, Strathmann J, et al. MicroRNA-132 dysregulation in schizophrenia has implications for both neurodevelopment and adult brain function. Proc Natl Acad Sci U S A 2012;109:3125-3130.

Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell 2002;110:689-699.

Moffat FL. Chapter 28: Clinical and pathologic prognostic and predictive factors. in Harris JR, Lippman ME, Morrow M, Osborne CK. Diseases of the Breast, 5th edition, Lippincott Williams & Wilkins, 2014.

Moldovan L, Batte KE, Trgovcich J, Wisler J, Marsh CB, Piper M. Methodological challenges in utilizing miRNAs as circulating biomarkers. J Cell Mol Med 2014;18:371-390.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nature Methods 2008;5:621-628.

Motawi TMK, Rizk SM, Shaker OG, Mokhtar OZH. MicroRNAs as predictor markers for response to interferon treatment of chronic hepatitis C genotype-4 in egyptian patients. PLoS ONE 2015;10:1-12.

Mu Y, Tang S, Sun W, Gao W, Wang M, Su X. Association of miR-193b down-regulation and miR-196a up-regulation with clinicopathological features and prognosis in gastric cancer. Asian Pac J Cancer Prev 2014;15:8893-8900.Wu W, Lin Z, Zhuang Z, Liang X. Expression profile of mammalian microRNAs in endometrioid adenocarcinoma. Eur J Cancer Prev 2009;18:50-55.

Myllykangas S, Buenrostro J, Ji HP. Overview of sequencing technology platforms. Bioinformatics for High Throughput Sequencing 2012:11.

Naderi A, Teschendorff AE, Barbosa-Morais N, Pinder SE, Green AR, Powe DG, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26:1507-1516.

Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. Microarray-based, high-throughput gene expression profiling of microRNAs. Nature Methods 2004;1:155-161.

Ng KW, Anderson C, Marshall EA, Minatel BC, Enfield KSS, Saprunoff HL, et al. Piwi-interacting RNAs in cancer: Emerging functions and clinical utility. Molecular Cancer 2016;15:1-13.

Nilsen TW. Mechanisms of microRNA-mediated gene regulation in animal cells. Trends in Genetics 2007;23:243-249.

Nygaard S, Jacobsen A, Lindow M, Eriksen J, Balslev E, Flyger H, et al. Identification and analysis of miRNAs in human breast cancer and teratoma samples using deep sequencing. BMC Medical Genomics 2009;2:35.

Oh SJ, Kim SM, Kim YO, Chang HK. Clinicopathologic implications of PIWIL2 expression in colorectal cancer. Korean journal of pathology 2012;46(4):318-323.

O'Hurley G, Sjosted E, Rahman A, Li B, Kampf C, Ponten F, et al. Garbage in, garbage out: A critical evaluation of strategies used for validation of immunohistochemical biomarkers. Molecular oncology 2014;8:783-798.

Okugawa Y, Toiyama Y, Toden S, Mitoma H, Nagasaka T, Tanaka K, et al. Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. Gut 2015.

Olvedy M, Scaravilli M, Hoogstrate Y, Visakorpi T, Jenster G, Martens-Uzunova E. A comprehensive repertoire of tRNA-derived fragments in prostate cancer. Oncotarget 2016.

Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res 2004;64:3087-3095.

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-2826.

Palakodeti D, Smielewska M, Lu YC, Yeo GW, Graveley BR. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. RNA 2008;14:1174-1186.

Palazzo AF, Gregory TR. The case for junk DNA. PLoS Genetics 2014;10:e1004351.

Pal-Bhadra M, Bhadra U, Birchler JA. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in drosophila. Mol Cell 2002;9:315-327.

Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA, et al. Heterochromatic silencing and HP1 localization in drosophila are dependent on the RNAi machinery. Science 2004;303:669-672.

Papagiannakopoulos T, Shapiro A, Kosik KS. MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. Cancer Res 2008;68:8164-8172.

Parker JS, Mullins M, Cheang M, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology 2009;27:1160-1167.

Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res 2009;37:7268-7280.

Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. RNA 2013;19:461-466.

Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. RNA 2013;19:461-6.

Peiró-Chova L, Peña-Chilet M, López-Guerrero JA, García-Giménez JL, Alonso-Yuste E, Burgues O, et al. High stability of microRNAs in tissue samples of compromised quality. Virchows archiv 2013;463:765-774.

Peng L, Song L, Liu C, Lv X, Li X, Jie J, et al. piR-55490 inhibits the growth of lung carcinoma by suppressing mTOR signaling. Tumour Biol 2015;37:2749-2756.

Pérez-Rivas LG, Jerez JM, Carmona R, de Luque V, Vicioso L, Claros MG, et al. A microRNA signature associated with early recurrence in breast cancer. PLoS ONE 2014;9:1-14.

Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast cancer research 2006;8:R23.

Pimentel F, Bonilla P, Ravishankar YG, Contag A, Gopal N, LaCour S, et al. Technology in MicroRNA profiling: Circulating MicroRNAs as noninvasive cancer biomarkers in breast cancer. J Lab Autom 2015;20:574-588.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.

Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: Approaches and considerations. Nature Reviews Genetics 2012;13:358-369.

Qiao D, Zeeman AM, Deng W, Looijenga LH, Lin HF. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. Oncogene 2002;21:3988-3999.

Rabinowits G, Gerçel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal MicroRNA: A diagnostic marker for lung cancer. Clin lung cancer 2009;10:42-46.

Radovich M, Clare SE, Atale R, Pardo I, Hancock BA, Solzak JP, et al. Characterizing the heterogeneity of triple-negative breast cancers using microdissected normal ductal epithelium and RNA-sequencing. Breast Cancer Res Treat 2014;143:57-68.

Raina M, Ibba M. tRNAs as regulators of biological processes. Front Genet 2014;5:171-171.

RajBhandary U, So¨ll D. tRNA : Structure, biosynthesis, and function. : Washington, D.C. : ASM Press, c1995; 1995.

Ramalingam P, Palanichamy JK, Singh A, Das P, Bhagat M, Kassab MA, et al. Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. RNA 2014;20:76-87.

Rani S, Gately K, Crown J, O'Byrne K, O'Driscoll L. Global analysis of serum microRNAs as potential biomarkers for lung adenocarcinoma. Cancer Biol Ther 2013;14:1104-1112.

Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nature Reviews Cancer 2004:309-314.

Raponi M, Dossey L, Jatkoe T, Wu X, Chen G, Fan H, et al. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. Cancer Res 2009;69:5776-5783.

Ravo M, Cordella A, Rinaldi A, Bruno G, Alexandrova E, Saggese P, et al. Small non-coding RNA deregulation in endometrial carcinogenesis. Oncotarget 2015;6:4677-4691.

Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Alvarado AS. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science 2005;310:1327-1330.

Reinhart BJ, Slack FJ. The 21-nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans. Nature 2000;403:901-906.

Renjie W, Haiqian L. MiR-132, miR-15a and miR-16 synergistically inhibit pituitary tumor cell proliferation, invasion and migration by targeting Sox5. Cancer Lett 2015;356:568-578.

Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 2006:873-875.

Ribelles N, Perez-Villa L, Jerez JM, Pajares B, Vicioso L, Jimenez B, et al. Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. Breast Cancer Res 2013;15:R98-R98.

Ring BZ, Seitz RS, Beck R, Shasteen WJ, Tarr SM, Cheang MC, et al. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3039-3047.

Ronchetti D, Mosca L, Cutrona G, Tuana G, Gentile M, Fabris S, et al. Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia. BMC Medical Genomics 2013;6:1-11.

Ronnau C, Verhaegh GW, Luna-Velez M, Schalken JA. Noncoding RNAs as novel biomarkers in prostate cancer. Biomed research international 2014;2014:591703.

Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al. MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol 2008;26:462-469.

Ross JS, Hatzis C, Symmans WF, Pusztai L, Hortobágyi G,N. Commercialized multigene predictors of clinical outcome for breast cancer. Oncologist 2008;13:477-493.

Ross JS. Multigene predictors in early-stage breast cancer: Moving in or moving out? Expert Review of Molecular Diagnostics 2008:129-135.

Rothé F, Ignatiadis M, Chaboteaux C, Haibe-Kains B, Kheddoumi N, Majjaj S, et al. Global MicroRNA expression profiling identifies MiR-210 associated with tumor proliferation, invasion and poor clinical outcome in breast cancer. PLoS ONE 2011;6:1-13.

Rouget C, Papin C, Boureux A, Meunier A, Franco B, Robine N, et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early drosophila embryo. Nature 2010;467:1128-1132.

Rouzier R, Pusztai L, Delaloge S, Gonzalez-Angulo A, Andre F, Hess KR, et al. Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. J Clin Oncol 2005;23:8331-8339.

Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, et al. Prader-willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat Genet 2008;40:719-721.

Sai Lakshmi S, Agrawal S. piRNABank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173-D177.

Sai Lakshmi S, Agrawal S. piRNABank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173-7.

Sanz-Pamplona R, Berenguer A, Cordero D, Mollev? DG, Crous-Bou M, Sole X, et al. Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. Molecular Cancer 2014;13:46.

Sapkota Y, Ghosh S, Lai R, Coe BP, Cass CE, Yasui Y, et al. Germline DNA copy number aberrations identified as potential prognostic factors for breast cancer recurrence. PLoS ONE 2013;8:1-11.

Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the drosophila melanogaster flamenco gene. Genetics 2004;166:1313-1321.

Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the drosophila melanogaster flamenco gene. Genetics 2004;166-1313-21.

Sasaki T, Shiohama A, Minoshima S, Shimizu N. Identification of eight members of the argonaute family in the human genome. Genomics 2003;82:323-330.

Sayed D, Abdellatif M. MicroRNAs in development and disease. Physiol Rev 2011;91:827-887.

Scarpati GDV, Calura E, Di Marino M, Romualdi C, Beltrame L, Malapelle U, et al. Analysis of differential miRNA expression in primary tumor and stroma of colorectal cancer patients. BioMed Research International 2014:1-8.

Schee K, Lorenz S, Worren MM, Günther C, Holden M, Hovig E, et al. Deep sequencing the MicroRNA transcriptome in colorectal cancer. PLoS ONE 2013;8:1-11.

Schummer M, Green A, Beatty JD, Karlan BY, Karlan S, Gross J, et al. Comparison of breast cancer to healthy control tissue discovers novel markers with potential for prognosis and early detection. PLoS ONE 2010;5:1-9.

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell 2003;115:199-208.

SEER Stat Fact Sheets: Female Breast Cancer; 2016 [Cited 2016 Jun]. Available from: http://seer.cancer.gov/statfacts/html/breast.html.

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao YJ, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395-399.

Sharma AK, Nelson MC, Brandt JE, Wessman M, Mahmud N, Weller KP, et al. Human CD34(+) stem cells express the hiwi gene, a human homologue of the drosophila gene piwi. Blood 2001;97:426-434.

Shen Y, Kong Z, Wan F, Wang H, Bian X, Gan H, et al. Downregulation of DAB2IP results in cell proliferation and invasion and contributes to unfavorable outcomes in bladder cancer. Cancer Science 2014;105:704-712.

Shigematsu M, Honda S, Kirino Y. Transfer RNA as a source of small functional RNA. J Mol Biol Mol Imaging 2014;1.

Shin VY, Ng EKO, Chan VW, Kwong A, Kent-Man Chu. A three-miRNA signature as promising non-invasive diagnostic marker for gastric cancer. Molecular Cancer 2015;14:1-9.

Shinden Y, Akiyoshi S, Ueo H, Nambara S, Saito T, Komatsu H, et al. Diminished expression of MiR-15a is an independent prognostic marker for breast cancer cases. Anticancer Res 2015;35:123-127.

Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: The vanguard of genome defence. Nature Reviews Molecular Cell Biology 2011;12:246-258.

Smith CM, Steitz JA. Sno storm in the nucleolus: New roles for myriad small RNPs. Cell 1997;89:669-672.

Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001(19):10869-10874.

Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003:8418-8423.

Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Raymond RS, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: Gene expression analyses across three different platforms. BMC Genomics 2006;7:127.

Sotiriou C, Author, Wirapati P, Author, Loi S, Author, Harris A, Author, Fox S, Author, Smeds J, Author, et al. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-272.

Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care? Nature Reviews Cancer 2007;7:545-553.

Sparano JA, Gray RJ, Makower DF, Pritchard K, Albain KS, Hayes DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. N Engl J Med 2015;373:2005-2014.

Spears JL, Rubio MAT, Sample PJ, Alfonzo JD. tRNA biogenesis and processing. RNA Metabolism in Trypanosomes 2012:99.

Sprinzl M, Vassilenko KS. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 2005;33:D139-D140.

Stanczyk J, Pedrioli DM, Brentano F, Sanchez-Pernaute O, Kolling C, Gay RE, et al. Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. Arthritis Rheum 2008;58:1001-1009.

Starega-Roslan J, Koscianska E, Kozlowski P, Krzyzosiak WJ. The role of the precursor structure in the biogenesis of microRNA. Cellular and molecular life sciences 2011;68:2859-2871.

Su H, Xu T, Ganapathy S, Shadfan M, Long M, Huang TH, et al. Elevated snoRNA biogenesis is essential in breast cancer. Oncogene 2014;33:1348-1358.

Su Y, Ni Z, Wang G, Cui J, Wei C, Wang J, et al. Aberrant expression of microRNAs in gastric cancer and biological significance of miR-574-3p. Int Immunopharmacol 2012;13:468-475.

Suh Y, Raulf N, Gaken J, Lawler K, Urbano TG, Bullenkamp J, et al. MicroRNA-196a promotes an oncogenic effect in head and neck cancer cells by suppressing annexin A1 and enhancing radioresistance. International Journal of Cancer 2015;137:1021-1034.

Sun GA, Wang YY, Sun LH, Luo H, Liu N, Fu Z, et al. Clinical significance of hiwi gene expression in gliomas. Brain Res 2011;1373:183-188.

Suzuki R, Honda S, Kirino Y. PIWI expression and function in cancer. Front Genet 2012;3:204-204.

Tafsiri E, Darbouy M, Shadmehr MB, Zagryazhskaya A, Alizadeh J, Karimipoor M. Expression of miRNAs in non-small-cell lung carcinomas and their association with clinicopathological features. Tumor Biol 2015;36:1603-1612.

Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res 2004;64:3753-3756.

Talbert PB, Henikoff S. Chromatin-based transcriptional punctuation. Genes Dev 2009;23:1037-1041.

Tam S, de Borja R, Tsao M, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Invest 2014;94:350-358.

Tang W, Zhu J, Su S, Wu W, Liu Q, Su F, et al. MiR-27 as a prognostic marker for breast cancer progression and patient survival. PLoS ONE 2012;7:1-7.

Tatarano S, Chiyomaru T, Kawakami K, Enokida H, Yoshino H, Hidaka H, et al. Novel oncogenic function of mesoderm development candidate 1 and its regulation by MiR-574-3p in bladder cancer cell lines. Int J Oncol 2012;40:951-959.

Taubert H, Greither T, Kaushal D, Würl P, Bache M, Bartel F, et al. Expression of the stem cell self-renewal gene hiwi and risk of tumour-related death in patients with soft-tissue sarcoma. Oncogene 2007;26:1098-1100.

Teixeira AL, Ferreira M, Silva J, Gomes M, Dias F, Santos JI, et al. Higher circulating expression levels of miR-221 associated with poor overall survival in renal cell carcinoma patients. Tumor Biol 2014;35:4057-4066.

Thomson T, Lin HF. The biogenesis and function of PIWI proteins and piRNAs: Progress and prospect. Annu Rev Cell Dev Biol 2009;25:355-376.

Troester MA, Lee MH, Carter M, Fan C, Cowan DW, Perez ER, et al. Activation of host wound responses in breast cancer microenvironment. Clin Cancer Res 2009;15:7020-7028.

Tu YK, Kellett M, Clerehugh V, Gilthorpe MS. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. Br Dent J 2005;199:457-61.

Ujihira T, Ikeda K, Suzuki T, Yamaga R, Sato W, Horie-Inoue K, et al. MicroRNA-574-3p, identified by microRNA library-based functional screening, modulates tamoxifen response in breast cancer. Sci Rep 2015;5:7641-7641.

Valleron W, Ysebaert L, Berquet L, Fataccioli V, Quelen C, Martin A, et al. Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. Blood 2012;120:3997-4005.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.

van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, et al. A signature pattern of stress-responsive MicroRNAs that can evoke cardiac hypertrophy and heart failure. Proc Natl Acad Sci U S A 2006; 103:18255-18260.

van't Veer L, Dai HY, van dV, He Y, Hart A, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-536.

Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: MicroRNAs can up-regulate translation. Science 2007;318:1931-1934.

Venkataraman S, Birks DK, Balakrishnan I, Alimova I, Harris PS, Patel PR, et al. MicroRNA 218 acts as a tumor suppressor by targeting multiple cancer phenotype-associated genes in medulloblastoma. J Biol Chem 2013;288:1918-1928.

Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010;28:1684-1691.

Volinia S, Calin GA, Liu C, Ambs S, Cimmino A, Petrocca F, et al. A MicroRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103:2257-2261.

Volinia S, Croce CM. Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. Proc Natl Acad Sci U S A 2013;110:7413-7417.

Voorhoeve PM, le Sage C, Schrier M, Gillis AJM, Stoop H, Nagel R, et al. Article: A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell 2006;124:1169-1181.

Wagner E, Liu Y, Schneider B, Storniolo AM, Han J, He C. Abstract 3268: Differences in expression of lncRNAs in breast tumor, adjacent normal-appearing breast tissue, and normal breast tissue from healthy donors. Cancer Research 2014;74:3268-3268.

Walter BA, Valera VA, Pinto PA, Merino MJ. Comprehensive microRNA profiling of prostate cancer. J Cancer 2013;4:350-357.

Wang B, Khachigian LM, Esau L, Birrer MJ, Zhao X, Parker MI, et al. A key role for early growth response-1 and nuclear factor-kappaB in mediating and maintaining GRO/CXCR2 proliferative signaling in esophageal cancer. Mol Cancer Res 2009;7:755-764.

Wang B, Komers R, Carew R, Winbanks CE, Xu B, Herman-Edelstein M, et al. Suppression of microRNA-29 expression by TGF-ß1 promotes collagen expression and renal fibrosis. J Am Soc Nephrol 2012;23:252-265.

Wang L, Zhu M, Ren A, Wu H, Han W, Tan R, et al. A ten-MicroRNA signature identified from a genome-wide MicroRNA expression profiling in human epithelial ovarian cancer. PLoS ONE 2014;9:1-8.

Wang Q, Han C, Milum K, Wani AA. Stem cell protein Piwil2 modulates chromatin modifications upon cisplatin treatment. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 2011;708:59-68.

Wang QR, Lee I, Ren JP, Ajay SS, Lee YS, Bao XY. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. Molecular therapy 2013;21:368-379.

Wang QR, Lee I, Ren JP, Ajay SS, Lee YS, Bao XY. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. Molecular therapy 2013;21:368-79.

Wang W, Liu L, Li G, Chen Y, Li C, Jin D, et al. Combined serum CA19-9 and miR-27a-3p in peripheral blood mononuclear cells to diagnose pancreatic cancer. Cancer Prev Res (Phila) 2013;6:331-338.

Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-679.

Wark AW, Lee HJ, Corn RM. Multiplexed detection methods for profiling MicroRNA expression in biological samples. Angewandte Chemie 2008;47:647-652.

Watanabe T, Lin H. Posttranscriptional regulation of gene expression by piwi proteins and piRNAs. Mol Cell 2014;56:18-27.

Wei R, Huang G, Zhang M, Li B, Zhang H, Shi M, et al. Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. Clin Cancer Res 2013;19:4780-4791.

Weigel MT. Current and emerging biomarkers in breast cancer: Prognosis and prediction. Endocr Relat Cancer 2010;17:R245-262.

Weinstein LB, Steitz JA. Guided tours: From precursor snoRNA to functional snoRNP. Curr Opin Cell Biol 1999;11:378-384.

Wei-Zhao Peng, Ma R, Wang F, Yu J, Zhi-Bin Liu. Role of miR-191/425 cluster in tumorigenesis and diagnosis of gastric cancer. International Journal of Molecular Sciences 2014;15:4031-4048.

Wek SA, Zhu S, Wek RC. The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. Mol Cell Biol 1995;15:4497-4506.

Wen-Kang Liu, Xiang-Yang Jiang, Zhen-Xi Zhang. Expression of PSCA, PIWIL1 and TBX2 and its correlation with HPV16 infection in formalin-fixed, paraffin-embedded cervical squamous cell carcinoma specimens. Arch Virol 2010;155:657-663.

Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 2004;116:855-862.

Wilson N. The miRNA registry. Nature Reviews Genetics 2004:244.

Winter AG, Sourvinos G, Allison SJ, Tosh K, Scott PH, Spandidos DA, et al. RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. Proc Natl Acad Sci U S A 2000;97:12619-12624.

Winter AG, Sourvinos G, Allison SJ, Tosh K, Scott PH, Spandidos DA, et al. RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. Proc Natl Acad Sci U S A 2000:12619.

Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009;11:228-234.

Wu X, Pu X, Zhao Z, Zhao Y, Kang X, Wu W, et al. The expression profiles of microRNAs in kaposi's sarcoma. Tumor Biol 2015:437.

Wu X, Pu X, Zhao Z, Zhao Y, Kang X, Wu W, et al. The expression profiles of microRNAs in kaposi's sarcoma. Tumor Biol 2015;36:437-446.

Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. RNA 2007;13:1668-1674.

Xiao G, Tang H, Wei W, Li J, Ji L, Ge J. Aberrant expression of MicroRNA-15a and MicroRNA-16 synergistically associates with tumor progression and prognosis in patients with colorectal cancer. Gastroenterology Research & Practice 2014:1-8.

Xiong J, Bing Z, Su Y, Deng D, Peng X. An integrated mRNA and microRNA expression signature for glioblastoma multiforme prognosis. PLoS ONE 2014;9:1-8.

Xu G, Yang F, Ding CL, Zhao LJ, Ren H, Zhao P, et al. Small nucleolar RNA 113-1 suppresses tumorigenesis in hepatocellular carcinoma. Molecular cancer 2014;13:216.

Xu W, Liu M, Peng X, Zhou P, Zhou J, Xu K, et al. miR-24-3p and miR-27a-3p promote cell proliferation in glioma cells via cooperative regulation of MXI1. Int J Oncol 2013;42:757-766.

Xu X, Wang W, Su N, Zhu X, Yao J, Gao W, et al. miR-374a promotes cell proliferation, migration and invasion by targeting SRCIN1 in gastric cancer. FEBS Lett 2015;589:407-413.

Yan H, Wu Q, Sun C, Ai L, Deng J, Zhang L, et al. piRNA-823 contributes to tumorigenesis by regulating de novo DNA methylation and angiogenesis in multiple myeloma. Leukemia (08876924) 2015;29:196-206.

Yan Ling JP, Abo R, Levine SS, Dedon PC. Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. Nucleic Acids Res 2014;42:e170.

Yan W, Li R, Liu Y, Yang P, Wang Z, Zhang C, et al. MicroRNA expression patterns in the malignant progression of gliomas and a 5-microRNA signature for prognosis. Oncotarget 2014;5:12908-12915.

Yang G, Han D, Chen X, Zhang D, Wang L, Shi C, et al. MiR-196a exerts its oncogenic effect in glioblastoma multiforme by inhibition of IκBα both in vitro and in vivo. Neuro Oncol 2014;16:652-661.

Yang X, Cheng YD, Lu Q, Wei JF, Yang HW, Gu M. Detection of stably expressed piRNAs in human blood. International journal of clinical and experimental medicine 2015;8(8):13353-13358.

Yang X, Cheng YD, Lu Q, Wei JF, Yang HW, Gu M. Detection of stably expressed piRNAs in human blood. International journal of clinical and experimental medicine 2015;8:13353-13358.

Yin H, Lin H. An epigenetic activation role of piwi and a piwi-associated piRNA in drosophila melanogaster. Nature 2007;450:304-308.

Yu L, Tumati V, Tseng S, Hsu F, Kim DN, Hong D, et al. DAB2IP regulates autophagy in prostate cancer in response to combined treatment of radiation and a DNA-PKcs inhibitor. Neoplasia 2012;14:1203-1212.

Yu S, Chen H, Chang G, Chen C, Chen H, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48-57.

Zamore PD. Somatic piRNA biogenesis. EMBO J 2010;29:3219-3221.

Zeng Y, Qu LK, Meng L, Liu CY, Dong B, Xing XF, et al. HIWI expression profile in cancer cells and its prognostic value for patients with colorectal cancer. Chin Med J 2011;124:2144-2149.

Zhang X, Azhar G, Wei JY. The expression of microRNA and microRNA clusters in the aging heart. PLoS ONE 2012;7:1-13.

Zhang X, Chen J, Radcliffe T, Lebrun DP, Tron VA, Feilotter H. An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. J Mol Diagn 2008;10:513-519.

Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 2005;436:214-220.

Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, et al. A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. Nucleic Acids Res 2015;43:10474-10491.

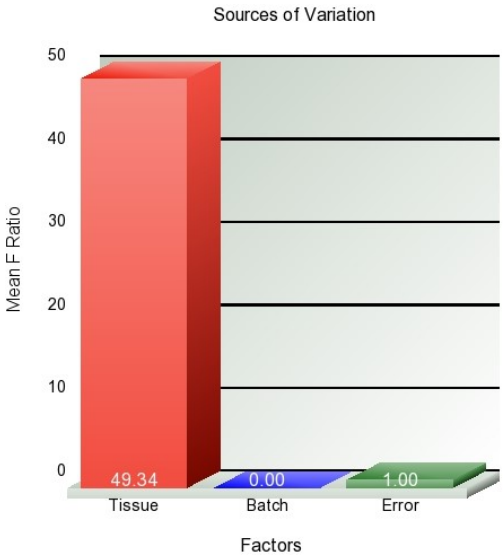
Zhu D, Zhu W, Fang C, Fan L, Zou Z, Wang Y, et al. miR-181a/b significantly enhances drug sensitivity in chronic lymphocytic leukemia cells via targeting multiple anti-apoptosis genes. Carcinogenesis 2012;33:1294-1301.

Zovoilis A, Agbemenyah HY, Agis-Balboa R, Stilling RM, Edbauer D, Rao P, et al. microRNA-34c is a novel target to treat dementias. EMBO J 2011;30:4299-4308.

9 Appendix

Figure 9.1 Batch effects correction using ANOVA model (miRNAs)

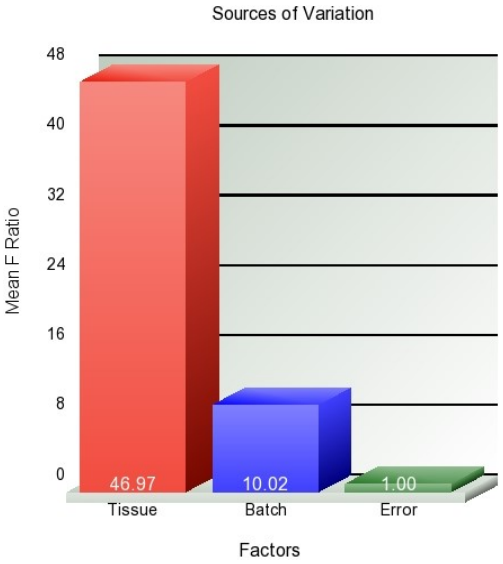
a. Data before batch correction



b. Data after batch correction

Figure 9.2 Batch effects correction using ANOVA model (piRNAs)

a. Data before batch correction



b. Data after batch correction

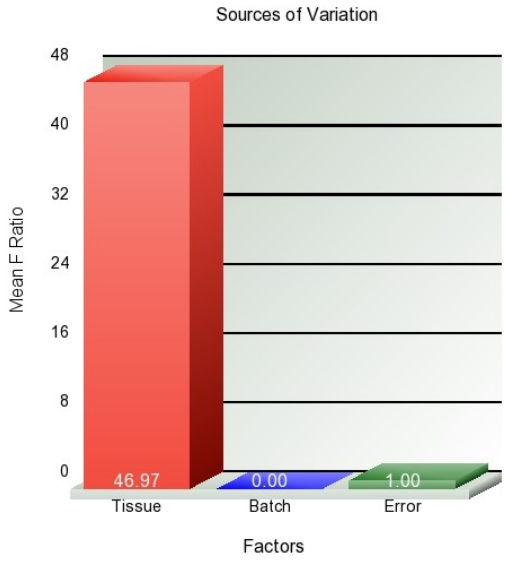


Figure 9.3 Batch effects correction using ANOVA model (tRNAs)

a. Data before batch correction **b. Data after batch correction**

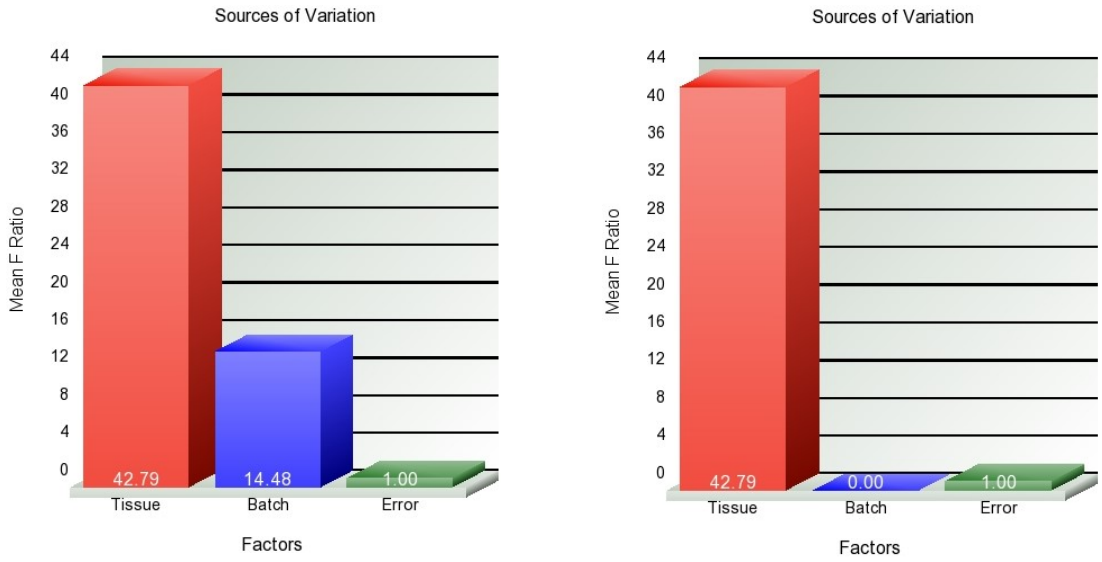
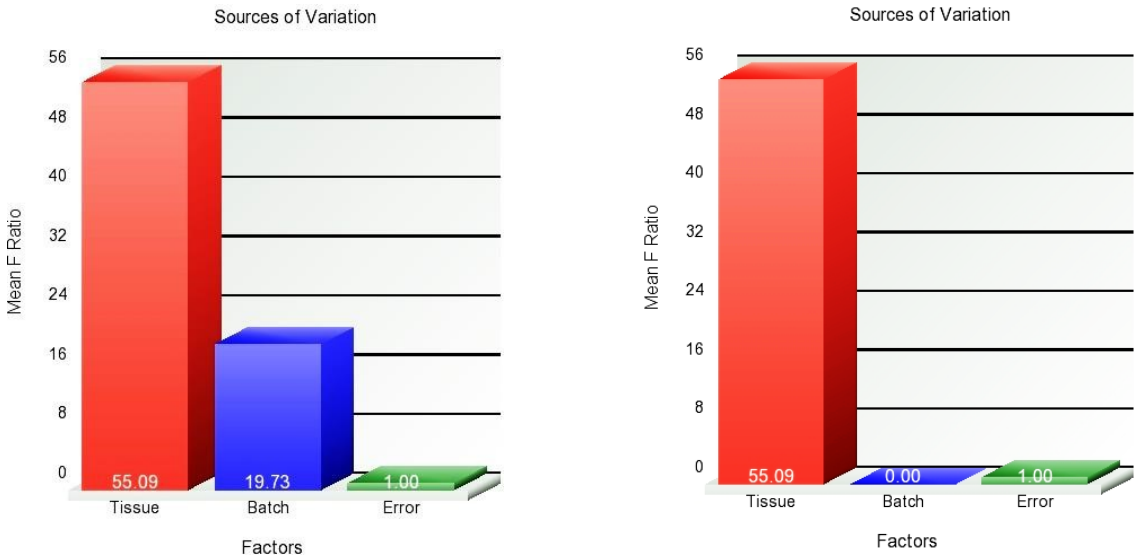


Figure 9.4 Batch effects correction using ANOVA model (snoRNAs)

a. Data before batch correction **b. Data after batch correction**



Tissue = Normal and Tumor tissue; Batch = Different batches in which the samples were sequenced

Appendix Figure 9.1-9.4: 104 tumor and 11 normal samples were sequenced in different batches. ANOVA model was used to capture the different sources of variation for miRNAs (Figure 9.1), piRNAs (Figure 9.2), tRNAs (Figure 9.3) and snoRNAs (Figure 9.4). All the factors having a mean F ratio above the mean F ratio of the error bar has to be corrected for (Figure a). Since normal and tumor tissues are sources of biological variation, the tissue factor was not corrected for and only batch (being a technical variation) was corrected for. The value of 0 for the factor batch in Figure b indicates that the data has been adjusted for batch effects.

Table 9.1 List of differentially expressed sncRNAs

sncRNA ID	p-value	FDR	Fold Change	Direction of Fold change
microRNAs				
hsa-let-7a-5p	2.32E-20	1.35E-19	-3.72	Down-regulated
hsa-let-7b-3p	2.00E-32	3.10E-31	-13.00	Down-regulated
hsa-let-7b-5p	9.92E-21	6.59E-20	-4.78	Down-regulated
hsa-let-7c-5p	8.16E-30	9.62E-29	-8.04	Down-regulated
hsa-let-7d-3p	1.04E-29	1.07E-28	-16.21	Down-regulated
hsa-let-7d-5p	9.82E-09	2.77E-08	-2.21	Down-regulated
hsa-let-7e-5p	1.98E-06	3.68E-06	-2.20	Down-regulated
hsa-let-7f-5p	1.35E-08	3.58E-08	5.96	Up-regulated
hsa-let-7i-5p	5.18E-05	8.03E-05	2.90	Up-regulated
hsa-miR-101-3p	2.71E-06	4.84E-06	5.01	Up-regulated
hsa-miR-103a-3p	9.08E-08	2.06E-07	6.93	Up-regulated
hsa-miR-107	1.76E-07	3.72E-07	10.86	Up-regulated
hsa-miR-10a-5p	0.0347755	0.0347755	2.42	Up-regulated
hsa-miR-125a-5p	5.30E-37	1.23E-35	-9.60	Down-regulated
hsa-miR-125b-5p	4.12E-29	3.83E-28	-4.91	Down-regulated
hsa-miR-126-3p	0.0008268	0.0010534	3.04	Up-regulated
hsa-miR-126-5p	0.0002196	0.0003094	12.23	Up-regulated
hsa-miR-141-3p	0.015057	0.0159125	4.03	Up-regulated
hsa-miR-142-5p	0.0175575	0.0183466	26.74	Up-regulated
hsa-miR-145-5p	3.05E-19	1.49E-18	-4.44	Down-regulated
hsa-miR-148a-3p	0.0014697	0.0017984	5.76	Up-regulated
hsa-miR-148a-5p	0.0044288	0.0050849	-2.01	Down-regulated
hsa-miR-150-5p	5.96E-09	1.73E-08	-3.81	Down-regulated
hsa-miR-151a-3p	0.0001104	0.0001604	2.99	Up-regulated
hsa-miR-155-5p	0.0304092	0.0310775	2.53	Up-regulated
hsa-miR-15a-5p	4.24E-08	1.04E-07	12.16	Up-regulated
hsa-miR-15b-5p	9.93E-18	4.40E-17	-2.92	Down-regulated
hsa-miR-16-5p	3.00E-11	1.12E-10	27.28	Up-regulated
hsa-miR-17-5p	0.0049192	0.0055791	3.58	Up-regulated
hsa-miR-181a-2-3p	6.09E-05	9.28E-05	-2.04	Down-regulated
hsa-miR-181a-5p	1.70E-09	5.27E-09	-2.36	Down-regulated

hsa-miR-181d-5p	0.0049893	0.0055905	3.80	Up-regulated
hsa-miR-182-5p	0.0287213	0.0296787	33.06	Up-regulated
hsa-miR-183-5p	0.0058226	0.0064464	9.67	Up-regulated
hsa-miR-192-5p	0.0031425	0.0037468	2.94	Up-regulated
hsa-miR-193b-3p	3.25E-15	1.37E-14	-4.34	Down-regulated
hsa-miR-195-5p	0.0036828	0.0042812	2.86	Up-regulated
hsa-miR-196b-5p	0.0133398	0.0142598	57.95	Up-regulated
hsa-miR-197-3p	6.00E-41	2.79E-39	-13.99	Down-regulated
hsa-miR-199a-3p	2.53E-05	4.15E-05	13.23	Up-regulated
hsa-miR-199a-5p	0.0003248	0.0004442	5.48	Up-regulated
hsa-miR-199b-3p	2.55E-05	4.15E-05	13.32	Up-regulated
hsa-miR-199b-5p	0.000889	0.0011173	3.65	Up-regulated
hsa-miR-19b-3p	1.12E-08	3.07E-08	3.36	Up-regulated
hsa-miR-200b-3p	0.0111238	0.0121707	3.75	Up-regulated
hsa-miR-205-5p	0.0004903	0.0006343	-2.49	Down-regulated
hsa-miR-20a-5p	0.0001847	0.0002642	17.93	Up-regulated
hsa-miR-21-3p	0.0025673	0.0031008	10.17	Up-regulated
hsa-miR-214-3p	4.35E-27	3.11E-26	-5.78	Down-regulated
hsa-miR-21-5p	5.84E-08	1.39E-07	3.87	Up-regulated
hsa-miR-22-3p	0.0013836	0.0017157	2.46	Up-regulated
hsa-miR-24-3p	3.50E-05	5.52E-05	2.43	Up-regulated
hsa-miR-26b-5p	8.99E-07	1.74E-06	32.23	Up-regulated
hsa-miR-27a-3p	6.42E-08	1.49E-07	6.46	Up-regulated
hsa-miR-27b-3p	2.23E-06	4.07E-06	4.35	Up-regulated
hsa-miR-28-3p	5.06E-06	8.71E-06	-2.05	Down-regulated
hsa-miR-28-5p	0.033187	0.0335477	2.18	Up-regulated
hsa-miR-29c-3p	7.34E-07	1.45E-06	2.90	Up-regulated
hsa-miR-30a-5p	0.0124884	0.0135049	12.89	Up-regulated
hsa-miR-30e-5p	1.23E-06	2.34E-06	24.39	Up-regulated
hsa-miR-320a	6.53E-45	6.07E-43	-7.08	Down-regulated
hsa-miR-335-5p	0.0004911	0.0006343	21.71	Up-regulated
hsa-miR-340-5p	1.11E-09	3.56E-09	2.52	Up-regulated
hsa-miR-342-3p	0.0003624	0.0004885	6.19	Up-regulated
hsa-miR-34a-5p	1.93E-05	3.27E-05	13.32	Up-regulated
hsa-miR-378a-3p	9.60E-05	0.0001417	-2.32	Down-regulated
hsa-miR-409-3p	3.47E-12	1.34E-11	-3.59	Down-regulated
hsa-miR-423-3p	1.79E-35	3.33E-34	-5.46	Down-regulated

hsa-miR-423-5p	9.69E-39	3.00E-37	-11.21	Down-regulated
hsa-miR-429	0.000259	0.0003595	2.46	Up-regulated
hsa-miR-486-5p	1.31E-20	8.13E-20	-15.22	Down-regulated
hsa-miR-497-5p	1.86E-07	3.84E-07	-2.78	Down-regulated
hsa-miR-574-3p	8.28E-30	9.62E-29	-5.80	Down-regulated
hsa-miR-654-3p	4.82E-10	1.66E-09	-3.64	Down-regulated
hsa-miR-660-5p	7.13E-07	1.44E-06	12.89	Up-regulated
hsa-miR-92a-3p	2.22E-28	1.88E-27	-5.33	Down-regulated
hsa-miR-92b-3p	8.75E-10	2.91E-09	-3.15	Down-regulated
hsa-miR-93-5p	0.003216	0.0037859	3.54	Up-regulated
hsa-miR-98-5p	1.11E-07	2.40E-07	22.15	Up-regulated
hsa-miR-99b-5p	1.36E-12	5.50E-12	-2.33	Down-regulated
piwi-interacting RNAs				
hsa_piR_008114	2.51E-31	6.29E-30	-8.54	Down-regulated
hsa_piR_019676	1.10E-28	9.13E-28	-8.01	Down-regulated
hsa_piR_000552	1.68E-30	2.10E-29	-6.56	Down-regulated
hsa_piR_020548	1.22E-27	7.62E-27	-4.90	Down-regulated
hsa_piR_008113	2.88E-20	1.44E-19	-3.71	Down-regulated
hsa_piR_016735	2.52E-14	9.01E-14	-2.88	Down-regulated
hsa_piR_020450	2.59E-11	7.20E-11	-2.74	Down-regulated
hsa_piR_017033	0.00085	0.001	-2.17	Down-regulated
hsa_piR_020365	0.026	0.026	2.23	Up-regulated
hsa_piR_019675	7.19E-09	1.80E-08	2.33	Up-regulated
hsa_piR_019914	0.007	0.007	2.42	Up-regulated
hsa_piR_015249	1.21E-13	3.77E-13	2.42	Up-regulated
hsa_piR_009294	3.78E-05	5.25E-05	2.95	Up-regulated
hsa_piR_021032	1.10E-05	1.72E-05	3.61	Up-regulated
hsa_piR_009051	1.32E-06	2.53E-06	4.38	Up-regulated
hsa_piR_000753	0.013	0.014	5.85	Up-regulated
hsa_piR_008112	1.50E-08	3.41E-08	6.94	Up-regulated
hsa_piR_020814	2.29E-06	4.09E-06	7.15	Up-regulated
hsa_piR_001318	3.55E-06	5.91E-06	8.71	Up-regulated
hsa_piR_006426	0.002	0.003	8.94	Up-regulated
hsa_piR_017184	3.76E-07	7.84E-07	9.17	Up-regulated
hsa_piR_020829	1.63E-05	2.39E-05	9.58	Up-regulated
hsa_piR_019912	1.56E-15	6.50E-15	16.64	Up-regulated

hsa_piR_018780	8.63E-05	0.0001	18.45	Up-regulated
hsa_piR_018849	0.023	0.024	27.26	Up-regulated
Transfer RNAs				
chr1.trna119-LysCTT	3.97E-15	1.67E-14	10.02	Up-regulated
chr1.trna16-HisGTG	0.00406	0.00454	100.75	Up-regulated
chr1.trna21-HisGTG	0.00482	0.00516	35.22	Up-regulated
chr1.trna26-AsnGTT	1.26E-10	2.45E-10	2.51	Up-regulated
chr1.trna34-LeuCAG	1.29E-14	4.91E-14	2.69	Up-regulated
chr1.trna36-LeuCAG	8.97E-17	1.70E-15	2.24	Up-regulated
chr1.trna38-LeuCAG	1.35E-15	9.14E-15	2.61	Up-regulated
chr1.trna40-LeuCAG	2.45E-15	1.17E-14	2.36	Up-regulated
chr1.trna42-LeuCAG	2.18E-15	1.14E-14	2.47	Up-regulated
chr1.trna47-AsnGTT	2.47E-07	3.61E-07	2.72	Up-regulated
chr1.trna54-LysTTT	8.15E-13	2.14E-12	3.64	Up-regulated
chr1.trna58-LeuCAA	5.54E-16	5.29E-15	2.04	Up-regulated
chr1.trna62-LysTTT	9.85E-13	2.42E-12	4.10	Up-regulated
chr1.trna67-LeuCAG	3.61E-15	1.61E-14	2.29	Up-regulated
chr1.trna7-AsnGTT	4.97E-09	9.00E-09	3.15	Up-regulated
chr1.trna9-ArgTCT	1.99E-07	2.96E-07	4.27	Up-regulated
chr2.trna27-GlyCCC	9.83E-13	2.42E-12	4.07	Up-regulated
chr5.trna11-LysCTT	1.09E-05	1.36E-05	7.08	Up-regulated
chr5.trna15-ValAAC	4.00E-09	7.42E-09	2.07	Up-regulated
chr5.trna9-LysCTT	1.10E-07	1.77E-07	15.79	Up-regulated
chr6.trna109-PheGAA	8.92E-06	1.17E-05	10.93	Up-regulated
chr6.trna114-ArgCCG	3.65E-07	5.14E-07	39.76	Up-regulated
chr6.trna129-MetCAT	7.54E-08	1.27E-07	11.06	Up-regulated
chr6.trna13-LysCTT	3.23E-18	8.82E-17	5.71	Up-regulated
chr6.trna142-MetCAT	3.48E-18	8.82E-17	6.26	Up-regulated
chr6.trna150-MetCAT	3.18E-18	8.82E-17	4.76	Up-regulated
chr6.trna169-MetCAT	0.00469	0.00509	33.56	Up-regulated
chr6.trna171-MetCAT	6.28E-15	2.51E-14	42.31	Up-regulated
chr6.trna1-GlnCTG	1.57E-15	9.18E-15	118.90	Up-regulated
chr6.trna33-HisGTG	1.36E-15	9.14E-15	2.54	Up-regulated
chr6.trna44-SerAGA	5.84E-05	7.04E-05	32.80	Up-regulated
chr6.trna45-AspGTC	1.15E-15	9.14E-15	4.06	Up-regulated
chr6.trna46-SerAGA	4.87E-16	5.29E-15	2.45	Up-regulated

chr6.trna48-AspGTC	1.44E-15	9.14E-15	9.65	Up-regulated
chr6.trna49-GlnCTG	0.00146	0.0017	2.83	Up-regulated
chr6.trna50-SerAGA	1.52E-12	3.51E-12	17.75	Up-regulated
chr6.trna51-SerTGA	8.54E-08	1.41E-07	2.84	Up-regulated
chr6.trna5-SerAGA	8.47E-07	1.15E-06	4.83	Up-regulated
chr6.trna72-PheGAA	5.86E-12	1.24E-11	3.63	Up-regulated
chr6.trna73-ArgCCG	9.96E-14	2.91E-13	2.47	Up-regulated
chr6.trna76-LysTTT	3.06E-07	4.39E-07	7.29	Up-regulated
chr6.trna7-LeuCAG	5.39E-09	9.52E-09	5215.67	Up-regulated
chr6.trna83-LeuTAA	1.12E-07	1.77E-07	13.69	Up-regulated
chr6.trna96-PheGAA	2.37E-10	4.50E-10	2.65	Up-regulated
chr8.trna10-MetCAT	6.25E-07	8.64E-07	21.58	Up-regulated
chr11.trna14-LysTTT	1.26E-16	1.91E-15	18.15	Up-regulated
chr11.trna17-ValTAC	0.0027	0.00306	2.52	Up-regulated
chr11.trna5-LysTTT	2.25E-15	1.14E-14	9.69	Up-regulated
chr12.trna11-PheGAA	1.49E-05	1.82E-05	9.57	Up-regulated
chr12.trna4-AspGTC	1.36E-12	3.24E-12	17.94	Up-regulated
chr12.trna5-AspGTC	1.70E-07	2.58E-07	18.72	Up-regulated
chr13.trna1-PheGAA	0.0019	0.00218	4.64	Up-regulated
chr13.trna3-GluTTC	0.00504	0.00533	8.98	Up-regulated
chr14.trna13-LysCTT	1.70E-12	3.80E-12	14.78	Up-regulated
chr15.trna11-GluTTC	0.01559	0.01623	4.77	Up-regulated
chr15.trna1-HisGTG	0.03092	0.03092	4.08	Up-regulated
chr15.trna2-LysCTT	0.02778	0.02815	4.66	Up-regulated
chr16.trna10-LysCTT	0.01619	0.01663	5.30	Up-regulated
chr16.trna17-LeuCAG	5.76E-12	1.24E-11	3.78	Up-regulated
chr16.trna1-ArgCCG	0.00447	0.00493	114.61	Up-regulated
chr16.trna22-MetCAT	2.46E-14	7.80E-14	2.31	Up-regulated
chr16.trna26-LeuCAG	1.01E-05	1.28E-05	12.05	Up-regulated
chr16.trna32-LysCTT	1.97E-14	7.06E-14	2.42	Up-regulated
chr16.trna34-GlyCCC	1.01E-05	1.28E-05	11.42	Up-regulated
chr16.trna7-LysCTT	4.80E-14	1.46E-13	2.94	Up-regulated
chr17.trna19-ArgTCG	1.42E-13	3.85E-13	2.56	Up-regulated
chr17.trna23-ArgCCG	2.04E-14	7.06E-14	2.61	Up-regulated
chr17.trna2-LysTTT	1.14E-13	3.20E-13	2.46	Up-regulated
chr17.trna3-GlnCTG	1.15E-06	1.54E-06	60.82	Up-regulated
chr17.trna4-ArgTCT	8.60E-05	0.0001	14.89	Up-regulated

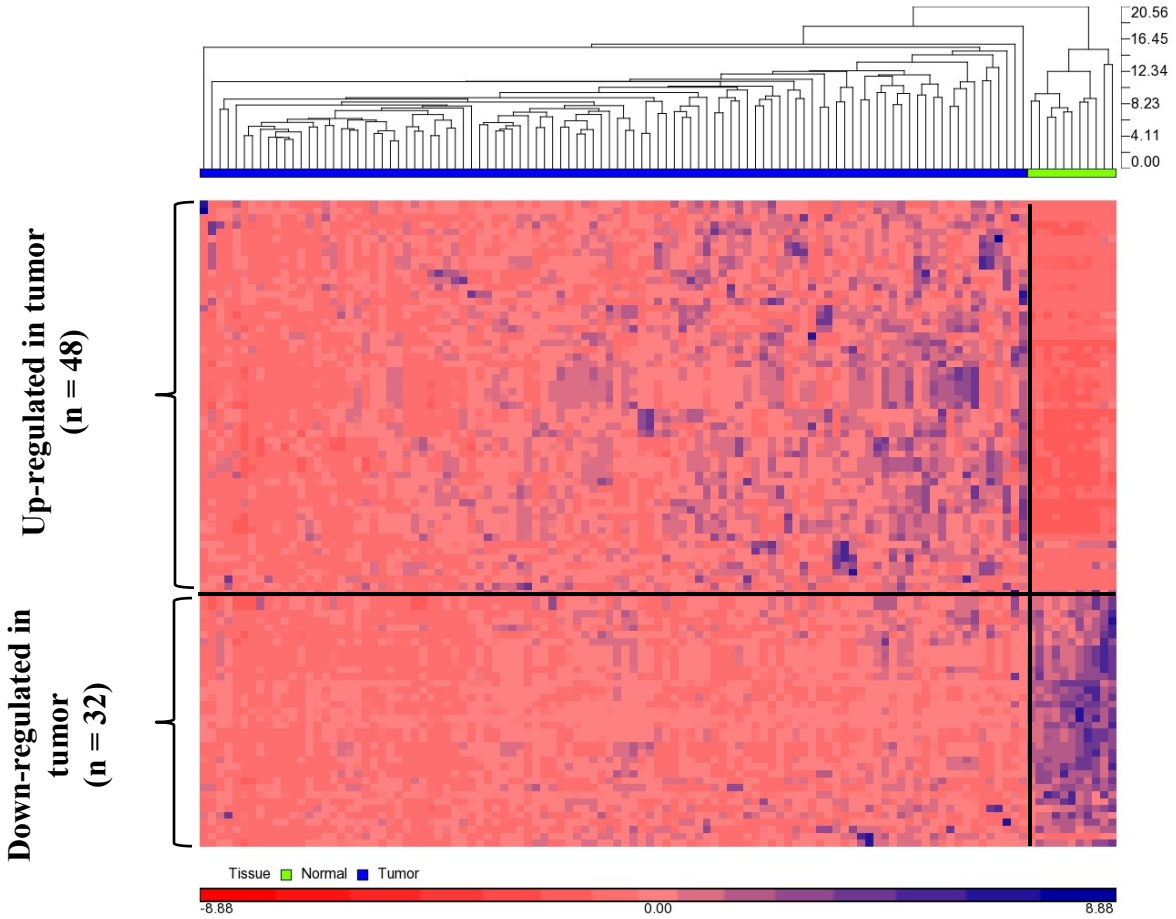
chr18.trna4-LysCTT	8.29E-12	1.66E-11	4.10	Up-regulated
chr19.trna14-PheGAA	5.56E-16	5.29E-15	2.49	Up-regulated
chr19.trna1-AsnGTT	2.43E-14	7.80E-14	3.98	Up-regulated
chr19.trna2-GlyTCC	1.44E-07	2.23E-07	12.94	Up-regulated
chr19.trna8-SeC(e)TCA	7.72E-12	1.59E-11	3.09	Up-regulated
chrX.trna4-ValTAC	3.54E-08	6.12E-08	17.17	Up-regulated
Small nucleolar RNAs				
SNORA45-201	4.98E-08	7.38E-08	3.92	Up-regulated
SNORA31-001	3.16E-06	4.08E-06	53.00	Up-regulated
SCARNA4-201	3.99E-06	4.99E-06	22.00	Up-regulated
SNORD101-201	7.24E-06	8.77E-06	23.99	Up-regulated
SNORA64-201	8.96E-05	9.95E-05	5.43	Up-regulated
SNORA48-201	0.000595231	0.000643493	2.90	Up-regulated
SNORD37-201	0.00134443	0.00141519	3.82	Up-regulated
SNORD1B-201	0.0017486	0.00179344	5.08	Up-regulated
SNORA65-201	0.00718696	0.00718696	3.24	Up-regulated
SNORD61-201	6.65E-36	2.05E-34	-7.90	Down-regulated
SNORD110-201	1.03E-35	2.05E-34	-24.22	Down-regulated
SNORD18A-201	2.33E-35	3.10E-34	-5.90	Down-regulated
SNORD68-201	4.13E-33	4.13E-32	-10.77	Down-regulated
SNORD46-201	1.02E-32	8.18E-32	-7.38	Down-regulated
SNORD89-201	6.12E-28	4.08E-27	-4.07	Down-regulated
SNORD59A-201	8.05E-28	4.60E-27	-4.36	Down-regulated
SNORD104-201	4.63E-27	2.32E-26	-4.37	Down-regulated
SNORD5-201	1.68E-26	7.46E-26	-7.17	Down-regulated
SNORD10-201	1.93E-26	7.71E-26	-4.06	Down-regulated
SNORD99-201	9.49E-21	3.45E-20	-2.97	Down-regulated
SNORA7B-201	1.69E-19	5.63E-19	-2.96	Down-regulated
SNORD20-201	3.72E-18	1.15E-17	-3.52	Down-regulated
SNORD119-201	4.44E-17	1.27E-16	-2.95	Down-regulated
SNORD13-201	1.17E-16	3.12E-16	-3.10	Down-regulated
SNORD102-201	6.37E-15	1.59E-14	-2.97	Down-regulated
SNORD34-201	1.21E-14	2.85E-14	-3.18	Down-regulated
SNORD63-201	2.95E-13	6.56E-13	-3.80	Down-regulated
SNORD58A-201	7.36E-13	1.55E-12	-2.90	Down-regulated
SNORD58B-201	1.59E-12	3.19E-12	-2.34	Down-regulated

SNORD12C-201	2.02E-12	3.86E-12	-3.32	Down-regulated
SNORD52-201	7.17E-12	1.30E-11	-3.95	Down-regulated
SNORD111B-201	5.98E-10	1.04E-09	-2.83	Down-regulated
SNORD15B-201	9.27E-10	1.54E-09	-2.83	Down-regulated
SNORD71-201	9.72E-10	1.56E-09	-2.10	Down-regulated
SNORA7A-201	2.65E-08	4.07E-08	-2.15	Down-regulated
SNORD42A-201	7.77E-08	1.11E-07	-2.40	Down-regulated
SNORD95-201	1.22E-07	1.69E-07	-2.43	Down-regulated
SNORD58C-201	3.06E-06	4.07E-06	-2.40	Down-regulated
SNORD84-201	2.00E-05	2.35E-05	-2.24	Down-regulated
SNORD17-201	2.26E-05	2.58E-05	-2.12	Down-regulated

FDR = False Discovery rate; Direction of effect indicates the direction of effect in tumor.

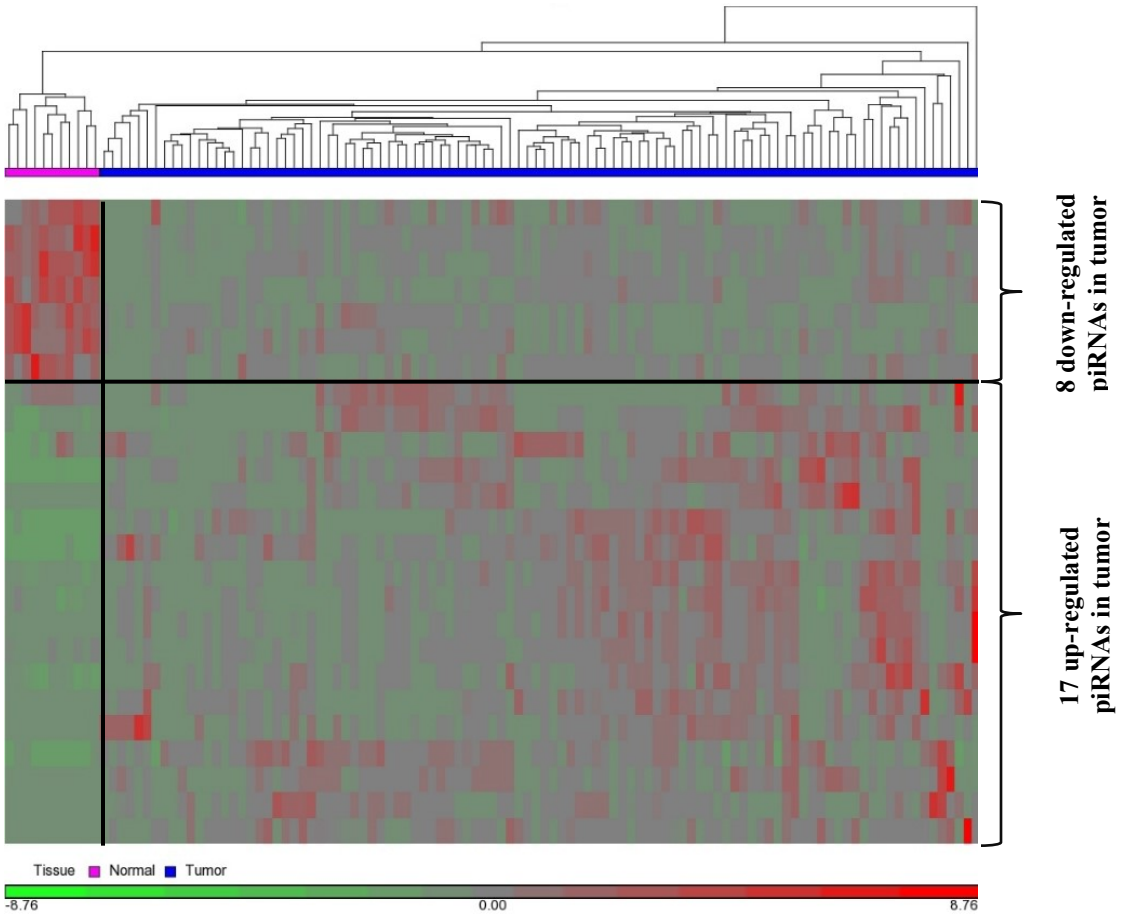
Appendix Table 9.1: 11 normal and 104 tumor samples were sequenced. sncRNAs were called differentially expressed if they exhibited fold change > 2.0 and FDR $\leq$ 0.05. 80 miRNAs, 25 piRNAs, 76 tRNAs and 40 snoRNAs were differentially expressed and are represented in this table.

Figure 9.5 Hierarchical clustering of differentially expressed miRNAs



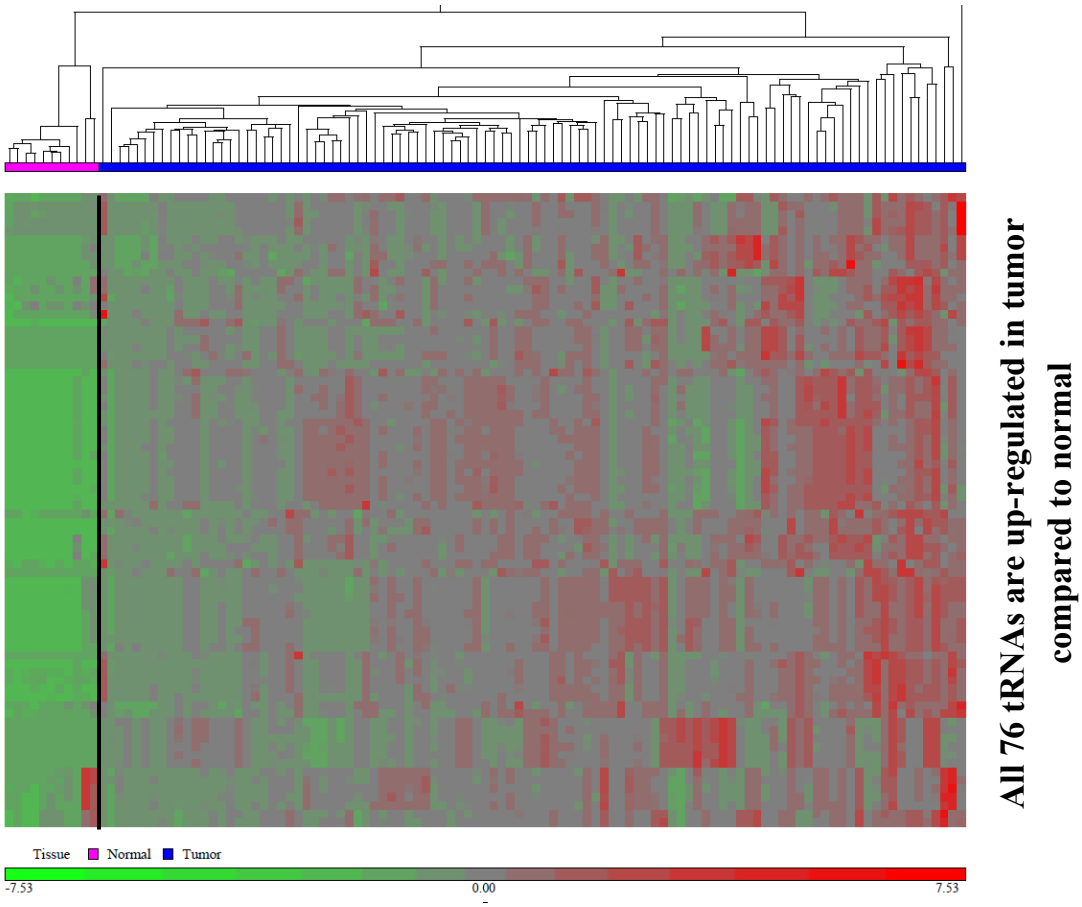
Appendix Figure 9.5: Unsupervised hierarchical clustering of 80 differentially expressed miRNAs was performed using Euclidean as distance measure and Average linkage method for linkage analysis. HC shows normal and tumor tissues as distinct clusters. 48 miRNAs were up-regulated in tumor and 32 miRNAs were down-regulated in tumor relative to normal tissues. Rows represent miRNAs and columns represent samples.

Figure 9.6 Hierarchical clustering of differentially expressed piRNAs



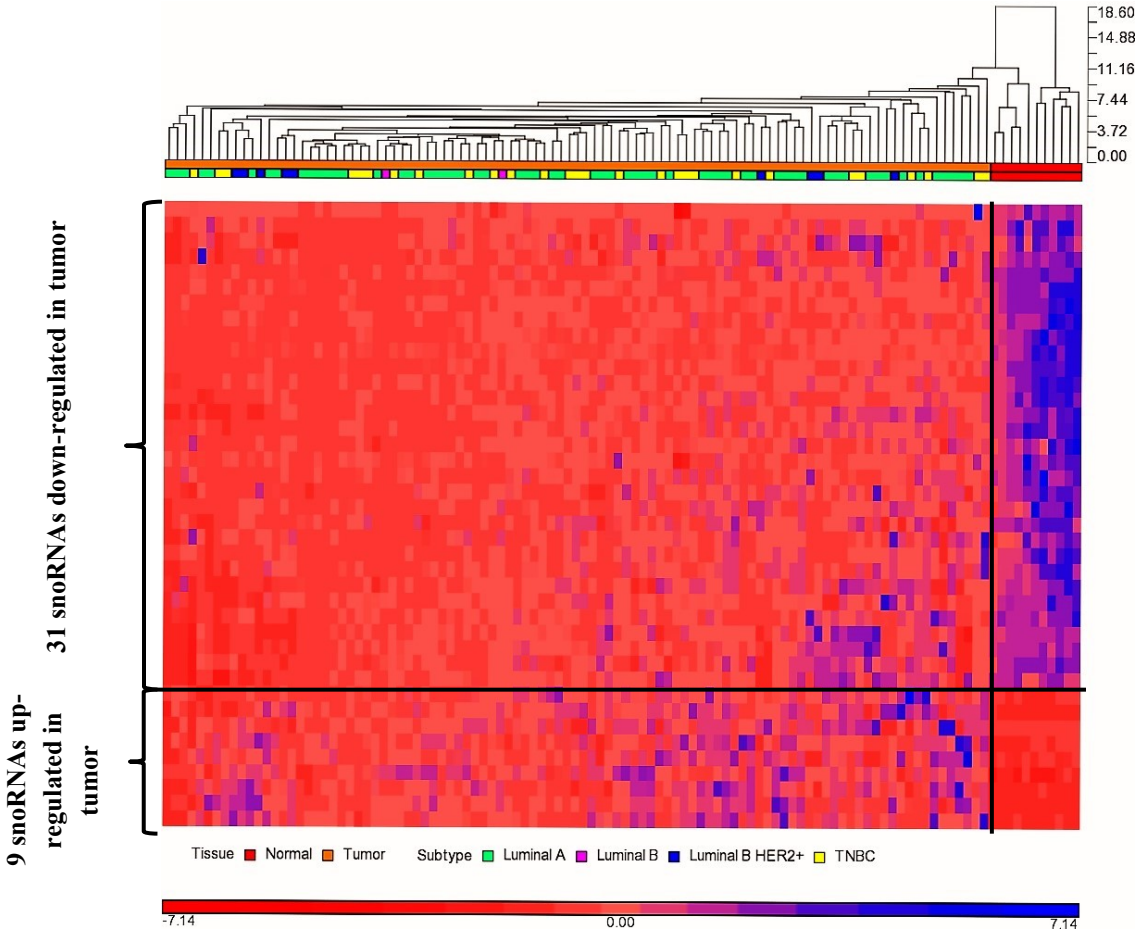
Appendix Figure 9.6: 25 differentially expressed piRNAs were used for unsupervised hierarchical clustering using average linkage method for linkage analysis and Euclidean distance measure.

Figure 9.7 Hierarchical clustering of differentially expressed tRNAs



Appendix Figure 9.7: 76 tRNAs were differentially expressed. The figure above represents the unsupervised hierarchical clustering of DE tRNAs. Euclidean distance was used as a measure of distance and average linkage method was used for linkage analysis. Samples are represented in columns and tRNAs are represented in rows. Blue bar indicates tumor samples. All 76 tRNAs are up-regulated in tumors compared to normal samples, as indicated by red and green colours, respectively.

Figure 9.8 Hierarchical clustering of differentially expressed snoRNAs



Appendix Figure 9.8: The 40 differentially expressed snoRNAs were subjected to unsupervised hierarchical clustering with average linkage and Euclidean as distance measure. The tumor samples (orange horizontal bar) were clearly separated from the normal samples (red horizontal bar).

Table 9.2 Fold changes of tRNA isoacceptors

Amino acid carried by the tRNA	Ala	Arg	Asn	Asp	Cys	Gln
Number of tRNAs profiled	42	33	36	20	32	34
Number of anticodons	3	5	2	1	2	2
Number of tRNAs retained after filtering	10	9	4	4	0	4
Number of anticodons seen in tRNAs retained after filtering	3	4	1	1		2
Fold changes of tRNA isoacceptors arising from different chromosomes	**AGC** - 1.86 (chr2), 1.15 (chr6), 1.65 (chr8)	**TCT** - 4.27 (chr1), 7.29 (chr17)	**GTT** - 2.51 (chr1), 2.65 (chr19), 2.72 (chr1), 3.15 (chr1)	**GTC** - 10.93 (chr12), 11.42 (chr6), 12.05 (chr6), 39.76 (chr12)		**CTG** - 2.47 (chr17), 2.94 (chr6), 3.78 (chr6)
	TGC - 1.14 (chr12), 1.17 (chr12), 1.16 (chr5), 1.14 (chr6)	**CCG** - 32.8 (chr16), 4.8 (chr17), 4.64 (chr6), 14.89 (chr6)				**TTG** - 1.84 (chr17)
	CGC - 1.12 (chr2, chr6)	**CCT** - 1.12 (chr16), 1.96 (chr7)				
		TCG - 2.84 (chr17)				

Lys	**Leu**	**Ile**	**His**	**Gly**	**Glu**
41	44	24	11	35	41
2	5	3	1	3	2
16	19	0	4	18	11
2	5		1	3	2
CTT - 10.02 (chr1), 6.26 (chr14), 42.3 (chr15), 118.9 (chr16), 9.64 (chr16), 17.75 (chr16), 5215.67 (chr18), 9.69 (chr5), 17.94 (chr5), 14.79 (chr6)	CAG - 2.69 (chr1), 2.44 (chr1), 2.61 (chr1), 2.36 (chr1), 2.47 (chr1), 2.29 (chr1), 2.54 (chr16), 2.45 (chr16), 2.49 (chr6)		**GTG** - 33.56 (chr15), 35.22 (chr1), 100.75 (chr1), 114.61 (chr6)	**CCC** - 1.81 (chr1), 2.52 (chr2), 2.83 (chr16)	**CTC** - 1.19 (chr1), 1.31 (chr1), 1.27 (chr1), 1.35 (chr1, chr6), 1.37 (chr1, chr6)
TTT - 3.64 (chr1), 4.1 (chr1), 4.07 (chr11), 2.07 (chr11), 3.63 (chr17), 4.1 (chr6)	CAA - 2.05 (chr1), 1.77 (chr6), 1.73 (chr6)			**GCC** - 2.49 (chr1), 2.57 (chr1), 2.59 (chr1), 2.55 (chr1), 1.88 (chr1), 1.79 (chr16), 1.89 (chr16), 1.96 (chr16), 1.84 (chr16, chr17), 1.99 (chr2, chr6), 2.81 (chr21)	**TTC** - 4.76 (chr15), 5.71 (chr13), 1.66 (chr13)
	AAG - 1.57 (chr14), 1.47 (chr16), 1.42 (chr5), 1.51 (chr5), 1.4 (chr6)			**TCC** - 21.58 (chr19)	
	TAG - 1.65 (chr16)				
	TAA - 3.97 (chr6)				

Val	Tyr	Trp	Thr	Ser	Pro	Phe	Met
37	16	8	22	27	25	15	20
3	1	1	3	5	3	1	1
22	0	2	0	6	0	6	11
3		1		2		1	1
CAC - 1.07 (chr6), 1.7 (chr1), 1.65 (chr1), 1.53 (chr1), 2.7 (chr19), 1.73 (chr5), 1.67 (chr5), 1.65 (chr5), 1.6 (chr5), 2.03 (chr6), 1.63 (chr6)		CCA - 1.04 (chr17), 1.76 (chr12)		TGA - 1.18 (chr10), 2.3 (chr6)		GAA - 15.79 (chr12), 11.06 (chr13), 13.69 (chr19), 18.72 (chr6), 60.82 (chr6), 12.94 (chr6)	CAT - 2.55 (chr1), 4.06 (chr16), 3.08 (chr17), 8.98 (chr6), 4.77 (chr6), 4.08 (chr6), 4.6 (chr6), 5.3 (chr6), 2.34 (chr6), 1.9 (chr6), 3.09 (chr8)
TAC - 7.08 (chr11), 17.17 (chrX)				AGA - 2.3 (chr6), 2.4 (chr6), 2.6 (chr6), 2.5 (chr6)			
AAC - 1.76 (chr3, chr5), 1.66 (chr5, chr6), 9.57 (chr5), 1.75 (chr5), 1.81 (chr6), 1.79 (chr6), 1.69 (chr6)							

Selenocysteine [SeC(e)]	3	1	2	1	TCA - 1.54 (chr22), 18.15 (chr19)				

Appendix Table 9.2: The isoacceptors identified for each amino acid and their corresponding fold changes are indicated. The origins of the isoacceptors on individual chromosomes are indicated in parenthesis.

Table 9.3 List of gene targets for piRNAs

Gene targets for piRNAs	Alignment score	Energy score	Gene name
List of gene targets for hsa_piR_009051			
NR5A2	176	-20.08	nuclear receptor subfamily 5, group A, member 2
ZNF177	175	-20.46	zinc finger protein 177
SSBP2	179	-23.62	single-stranded DNA binding protein 2
KCNS1	170	-23.96	Potassium Voltage-Gated Channel, Delayed-Rectifier, Subfamily S, Member 1
ZNF765	171	-24.68	zinc finger protein 765
TMCC3	170	-25.63	transmembrane and coiled-coil domain family 3
FOXO4	177	-26.89	forkhead box O4
TRIM2	173	-28.11	tripartite motif containing 2
SCARA3	178	-28.79	scavenger receptor class A, member 3
KCNA1	171	-32.74	potassium voltage-gated channel, shaker-related subfamily, member 1 (episodic ataxia with myokymia)
List of gene targets for hsa_piR_021032			
COL4A3	170	-21.43	collagen, type IV, alpha 3 (Goodpasture antigen)
SOBP	170	-21.79	sine oculis binding protein homolog (Drosophila)
ZBTB33	170	-21.86	zinc finger and BTB domain containing 33
PKN2	170	-22.46	protein kinase N2
ADHFE1	170	-22.62	alcohol dehydrogenase, iron containing, 1
IRX4	170	-23.67	iroquois homeobox 4
GTF2IRD2	170	-24.04	GTF2I repeat domain containing 2
RNF213	170	-25.02	ring finger protein 213
EPDR1	170	-31.49	ependymin related 1
RNF180	171	-20.16	ring finger protein 180
ZNF333	171	-21.6	zinc finger protein 333
ADAM33	171	-21.73	ADAM metallopeptidase domain 33
FBXO9	171	-21.73	F-box protein 9
LRRC28	171	-21.73	leucine rich repeat containing 28
SEMA5A	171	-21.73	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A

ZNF506	171	-21.73	zinc finger protein 506
CIRBP	171	-22.38	cold inducible RNA binding protein
SMAD9	171	-22.51	SMAD family member 9
FRMD4A	171	-23.34	FERM domain containing 4A
SYNPO2	171	-23.55	synaptopodin 2
ARHGAP26	171	-23.56	Rho GTPase activating protein 26
CTDSPL	171	-23.56	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like
SLC25A26	171	-23.56	solute carrier family 25 (S-adenosylmethionine carrier), member 26
VLDLR	171	-23.77	very low density lipoprotein receptor
MAPKBP1	171	-25.57	mitogen-activated protein kinase binding protein 1
PTCHD1	171	-25.57	patched domain containing 1
TPT1	171	-26.34	tumor protein, translationally-controlled 1
LDB3	171	-26.82	LIM domain binding 3
MYOZ3	171	-27.33	myozenin 3
CRTAP	171	-27.67	cartilage associated protein
GNAI1	171	-29.84	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1
FGD5	172	-21.2	FYVE, RhoGEF and PH domain containing 5
KIRREL	172	-22.49	kin of IRRE like (Drosophila)
TGFB1I1	172	-22.63	transforming growth factor beta 1 induced transcript 1
LTBP2	172	-24.17	latent transforming growth factor beta binding protein 2
SV2B	172	-24.82	synaptic vesicle glycoprotein 2B
ZNF488	172	-25.19	zinc finger protein 488
KIF1B	172	-25.49	kinesin family member 1B
TLE4	172	-26.4	transducin-like enhancer of split 4
PER2	172	-26.64	period circadian clock 2
KLHL3	172	-28.05	kelch-like family member 3
TMEM64	173	-23.15	transmembrane protein 64
VSX1	173	-24.08	visual system homeobox 1
TRIOBP	173	-24.09	TRIO and F-actin binding protein
ST8SIA3	173	-28.16	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 3
PPARA	174	-23.49	peroxisome proliferator-activated receptor alpha

TRPM3	174	-23.78	transient receptor potential cation channel, subfamily M, member 3
DNAJC18	174	-24.29	DnaJ (Hsp40) homolog, subfamily C, member 18
PPP1R12B	174	-24.43	protein phosphatase 1, regulatory subunit 12B
MYBPC1	174	-25.16	myosin binding protein C, slow type
NPTX1	174	-25.18	neuronal pentraxin I
FGFR1	174	-25.55	fibroblast growth factor receptor 1
SLC16A12	174	-26.16	solute carrier family 16, member 12
NBEA	174	-28.54	neurobeachin
ALX4	174	-29.2	ALX homeobox 4
TNRC6B	175	-22.28	trinucleotide repeat containing 6B
FOSB	175	-25.06	FBJ murine osteosarcoma viral oncogene homolog B
LEPR	175	-25.14	leptin receptor
BVES	175	-25.47	**BVES** antisense RNA 1
PIWIL2	175	-25.7	piwi-like RNA-mediated gene silencing 2
SHROOM4	175	-26.35	shroom family member 4
MSRB3	175	-26.89	methionine sulfoxide reductase B3
CLDN19	175	-27.23	claudin 19
ICA1L	175	-27.23	islet cell autoantigen 1,69kDa-like
KLK10	175	-27.46	kallikrein-related peptidase 10
METTL7A	175	-27.46	methyltransferase like 7A
PDE7B	175	-27.46	phosphodiesterase 7B
MRAS	175	-27.62	muscle RAS oncogene homolog
ART4	175	-27.97	ADP-ribosyltransferase 4 (Dombrock blood group)
CAV2	175	-27.97	caveolin 2
MAN2A2	175	-28.33	mannosidase, alpha, class 2A, member 2
TXNIP	175	-31.5	thioredoxin interacting protein
CFL2	176	-24.27	cofilin 2 (muscle)
NALCN	177	-27.04	sodium leak channel, non-selective
FGD4	178	-26.18	FYVE, RhoGEF and PH domain containing 4
RPL13	178	-27.15	ribosomal protein L13
HLF	178	-27.64	hepatic leukemia factor
KCNMA1	179	-21.93	potassium large conductance calcium-activated channel, subfamily M, alpha member 1
POLR1E	179	-25.26	polymerase (RNA) I polypeptide E, 53kDa
AKAP9	179	-25.79	A kinase (PRKA) anchor protein 9
FANCC	179	-25.79	Fanconi anemia, complementation group C

SETD5	179	-26.06	SET domain containing 5
NEDD9	179	-26.79	neural precursor cell expressed, developmentally down-regulated 9
ATP1A2	179	-27.15	ATPase, Na+/K+ transporting, alpha 2 polypeptide
AP3B2	179	-27.15	adaptor-related protein complex 3, beta 2 subunit
CEP68	179	-27.15	centrosomal protein 68kDa
CCBE1	179	-27.15	collagen and calcium binding EGF domains 1
DIXDC1	179	-27.15	DIX domain containing 1
CPM	179	-27.15	carboxypeptidase M
FGFR2	179	-27.15	fibroblast growth factor receptor 2
KRT15	179	-27.15	keratin 15
MICAL3	179	-27.15	microtubule associated monooxygenase, calponin and LIM domain containing 3
PDCD4	179	-27.15	programmed cell death 4 (neoplastic transformation inhibitor)
NRIP2	179	-27.15	nuclear receptor interacting protein 2
PCGF5	179	-27.15	polycomb group ring finger 5
PARK2	179	-27.15	parkin RBR E3 ubiquitin protein ligase
NRP1	179	-27.15	neuropilin 1
SLC26A4	179	-27.15	solute carrier family 26 (anion exchanger), member 4
SLC25A37	179	-27.15	solute carrier family 25 (mitochondrial iron transporter), member 37
STAC2	179	-27.15	SH3 and cysteine rich domain 2
ST13	179	-27.15	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)
ZNF667	179	-27.15	zinc finger protein 667
UACA	179	-27.15	uveal autoantigen with coiled-coil domains and ankyrin repeats
ZNF483	179	-27.15	zinc finger protein 483
PKD2	179	-28.77	polycystic kidney disease 2 (autosomal dominant)
GAB1	179	-28.9	GRB2-associated binding protein 1
CLPX	179	-28.98	caseinolytic mitochondrial matrix peptidase chaperone subunit
CHNRB1	179	-28.98	
CD160	179	-28.98	**CD160** molecule
CCDC69	179	-28.98	coiled-coil domain containing 69
ESR2	179	-28.98	estrogen receptor 2 (ER beta)

HB3ST1	179	-28.98	
GSTM3	179	-28.98	glutathione S-transferase mu 3 (brain)
GLRA3	179	-28.98	glycine receptor, alpha 3
GNAL	179	-28.98	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type
MCAM	179	-28.98	melanoma cell adhesion molecule
LEP	179	-28.98	leptin
LYRM7	179	-28.98	LYR motif containing 7
OTUD6A	179	-28.98	OTU deubiquitinase 6A
NMT2	179	-28.98	N-myristoyltransferase 2
RNF157	179	-28.98	ring finger protein 157
SLC14A2	179	-28.98	solute carrier family 14 (urea transporter), member 2
SCN2B	179	-28.98	sodium channel, voltage-gated, type II, beta subunit
ROBO4	179	-28.98	roundabout, axon guidance receptor, homolog 4 (Drosophila)
SOX7	179	-28.98	SRY (sex determining region Y)-box 7
SMYD4	179	-28.98	SET and MYND domain containing 4
SEMA3E	179	-28.98	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3E
SAMD5	179	-28.98	sterile alpha motif domain containing 5
RPS9	179	-28.98	ribosomal protein S9
SH3TC2	179	-28.98	SH3 domain and tetratricopeptide repeats 2
SVEP1	179	-28.98	sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1
TBRG1	179	-28.98	transforming growth factor beta regulator 1
TCF7	179	-28.98	transcription factor 7 (T-cell specific, HMG-box)
TRPM6	179	-28.98	transient receptor potential cation channel, subfamily M, member 6
TMEM130	179	-28.98	transmembrane protein 130
ZNF154	179	-28.98	zinc finger protein 154
SLC2A4	179	-29.07	solute carrier family 2 (facilitated glucose transporter), member 4
LIAS	179	-29.08	lipoic acid synthetase
RASSF6	179	-29.08	Ras association (RalGDS/AF-6) domain family member 6
FILIP1	179	-29.57	filamin A interacting protein 1
ANXA11	179	-30.15	annexin A11

MON1B	179	-33.09	MON1 secretory trafficking family member B
ZNF366	179	-33.09	zinc finger protein 366
ZNF135	180	-27.12	zinc finger protein 135
USHBP1	180	-27.82	Usher syndrome 1C binding protein 1
BICC1	182	-24.7	BicC family RNA binding protein 1
ACAT1	183	-25.7	acetyl-CoA acetyltransferase 1
PCDH11Y	183	-26.05	protocadherin 11 Y-linked
PAK3	183	-26.05	p21 protein (Cdc42/Rac)-activated kinase 3
FUT6	183	-27.89	fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
IDS	183	-27.89	iduronate 2-sulfatase
KCNN1	183	-27.89	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 1
LRRC2	183	-27.89	leucine rich repeat containing 2
LYRM4	183	-27.89	LYR motif containing 4
SLC14A1	183	-27.89	solute carrier family 14 (urea transporter), member 1 (Kidd blood group)
ZHX3	183	-27.89	zinc fingers and homeoboxes 3
ZNF561	183	-27.89	zinc finger protein 561
XPNPEP3	183	-28.12	X-prolyl aminopeptidase (aminopeptidase P) 3, putative
CA5B	183	-28.89	carbonic anhydrase VB, mitochondrial
KCNMB1	183	-29.49	potassium large conductance calcium-activated channel, subfamily M, beta member 1
CHDH	183	-31.6	choline dehydrogenase
SLC22A3	184	-25.9	solute carrier family 22 (organic cation transporter), member 3
GGT6	184	-26.33	gamma-glutamyltransferase 6
PLXDC1	184	-31.84	plexin domain containing 1
CUBN	184	-32.53	cubilin (intrinsic factor-cobalamin receptor)
FBXL2	185	-26.74	F-box and leucine-rich repeat protein 2
KCNH8	185	-28.44	potassium voltage-gated channel, subfamily H (eag-related), member 8
BHMT2	185	-30.57	betaine--homocysteine S-methyltransferase 2
FER	185	-36.78	**fer** (fps/fes related) tyrosine kinase
NRXN1	186	-25.31	neurexin 1
NFAT5	187	-29.16	nuclear factor of activated T-cells 5, tonicity-responsive
PARD6G	187	-32.32	par-6 family cell polarity regulator gamma

DENND2A	189	-30.46	DENN/MADD domain containing 2A
TCEAL7	189	-31.22	transcription elongation factor A (SII)-like 7
NOSTRIN	190	-31.24	nitric oxide synthase trafficking
SHF	191	-30.12	Src homology 2 domain containing F
EPB42	191	-32.31	erythrocyte membrane protein band 4.2
ZNF589	191	-32.31	zinc finger protein 589
ITIH5	192	-31.93	inter-alpha-trypsin inhibitor heavy chain family, member 5
TNFSF12	197	-39.56	tumor necrosis factor (ligand) superfamily, member 12
List of gene targets for hsa_piR_015249			
FOXP2	174	-21.46	forkhead box P2
List of gene targets for hsa_piR_004153			
IL22RA1	170	-28.7	Interleukin 22 Receptor, Alpha 1
UST	170	-30.51	uronyl-2-sulfotransferase
FOXP2	170	-31.14	forkhead box P2
RPS9	170	-31.78	ribosomal protein S9
LILRB5	170	-39.16	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 5
AFP	171	-26.72	alpha-fetoprotein
MYOZ3	171	-27.7	myozenin 3
ARHGAP19	171	-31.16	Rho GTPase activating protein 19
APOL4	171	-31.74	apolipoprotein L, 4
FRMD1	171	-36.5	FERM domain containing 1
AGPAT2	171	-39.15	1-acylglycerol-3-phosphate O-acyltransferase 2
MLXIPL	171	-40.44	MLX interacting protein-like
PLEKHA4	171	-45.37	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 4
GRID1	172	-29.15	glutamate receptor, ionotropic, delta 1
SCRT1	172	-31.47	scratch family zinc finger 1
GRPR3	172	-31.73	gastrin-releasing peptide receptor
KCNMB1	172	-35.46	potassium large conductance calcium-activated channel, subfamily M, beta member 1
AGBL5	172	-36.39	ATP/GTP binding protein-like 5
DAB2IP	172	-41.26	DAB2 interacting protein
CCDC38	173	-25.29	coiled-coil domain containing 38
CTNND1	173	-32.75	catenin (cadherin-associated protein), delta 1
CPN2	173	-33.79	carboxypeptidase N, polypeptide 2

ZNF395	173	-34.07	zinc finger protein 395
TMCC3	173	-40.19	transmembrane and coiled-coil domain family 3
ITSN1	174	-33.86	intersectin 1 (SH3 domain protein)
CRHR1	174	-33.99	corticotropin releasing hormone receptor 1
GLB1L3	174	-36.2	galactosidase, beta 1-like 3
MGLL	174	-36.67	monoglyceride lipase
CPEB1	174	-38.76	cytoplasmic polyadenylation element binding protein 1
LAMA4	174	-45.36	laminin, alpha 4
CALCR	175	-34.27	calcitonin receptor
CYGB	175	-34.59	cytoglobin
ST8SIA3	175	-36.46	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 3
CLEC4F	175	-43.15	C-type lectin domain family 4, member F
PIGR	176	-38.81	polymeric immunoglobulin receptor
OLFML2A	178	-37.25	olfactomedin-like 2A
LUZP1	178	-39.93	leucine zipper protein 1
SHANK3	178	-45.21	SH3 and multiple ankyrin repeat domains 3
ADAM11	181	-42.41	ADAM metallopeptidase domain 11
CAV1	185	-38.06	caveolin 1, caveolae protein, 22kDa
H6PD	188	-40.62	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)
ALPL	193	-36.67	alkaline phosphatase, liver/bone/kidney
List of gene targets for hsa_piR_017716			
CLIP3	170	-27.29	CAP-GLY domain containing linker protein 3
ERICH1	170	-27.51	glutamate-rich 1
ACSM	170	-28.04	acyl-CoA synthetase medium-chain family member 2B
TIFA	170	-30.23	TRAF-interacting protein with forkhead-associated domain
SLC34A2	170	-31.76	solute carrier family 34 (type II sodium/phosphate contransporter), member 2
PIK3IP1	170	-32.25	phosphoinositide-3-kinase interacting protein 1
EPHA2	170	-32.49	EPH receptor A2
INMT	170	-33.15	indolethylamine N-methyltransferase
SCARA3	170	-34.01	scavenger receptor class A, member 3
ALPL	170	-34.84	alkaline phosphatase, liver/bone/kidney
TNFSF12	170	-40.14	tumor necrosis factor (ligand) superfamily, member 12

FIBD1	171	-25.2	
NEIL1	171	-32.19	nei endonuclease VIII-like 1 (E. coli)
GREM2	171	-33.03	gremlin 2, DAN family BMP antagonist
SIRPA	171	-33.84	signal-regulatory protein alpha
IRX4	171	-34.5	iroquois homeobox 4
RAPGEF3	171	-34.88	Rap guanine nucleotide exchange factor (GEF) 3
TXNRD2	171	-35.01	thioredoxin reductase 2
CCDC120	171	-37.03	coiled-coil domain containing 120
EPAS1	171	-38.36	endothelial PAS domain protein 1
GNAL	172	-28.98	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type
GNG2	172	-30.1	guanine nucleotide binding protein (G protein), gamma 2
P2RX3	172	-31.03	purinergic receptor P2X, ligand-gated ion channel, 3
KLHL29	172	-31.19	kelch-like family member 29
ACVR1C	172	-31.84	activin A receptor, type IC
ACACB	172	-32.16	acetyl-CoA carboxylase beta
LMOD3	172	-33.14	leiomodin 3 (fetal)
RGL1	172	-35.29	ral guanine nucleotide dissociation stimulator-like 1
KCNK17	172	-37.55	potassium channel, subfamily K, member 17
LYRM4	172	-40.51	LYR motif containing 4
ANGPTL4	173	-28.16	angiopoietin-like 4
GRID1	173	-28.17	glutamate receptor, ionotropic, delta 1
IL33	173	-32.77	interleukin 33
TLN2	173	-33.7	talin 2
CLEC4M	173	-34.43	C-type lectin domain family 4, member M
ADAM11	173	-36.11	ADAM metallopeptidase domain 11
TRIM2	174	-31.87	tripartite motif containing 2
KCNMB1	174	-32.43	potassium large conductance calcium-activated channel, subfamily M, beta member 1
SLC23A2	174	-32.64	solute carrier family 23 (ascorbic acid transporter), member 2
JAM3	174	-32.77	junctional adhesion molecule 3
ST8SIA2	174	-33.46	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2
LAMA4	174	-44.11	laminin, alpha 4
FUNDC2	175	-28.18	FUN14 domain containing 2

IL22RA1	175	-29.87	interleukin 22 receptor, alpha 1
FOXP2	175	-32.31	forkhead box P2
RPS9	175	-32.95	ribosomal protein S9
PPARA	175	-34.38	peroxisome proliferator-activated receptor alpha
RSPO1	175	-36.28	R-spondin 1
CX3CL1	175	-37.43	chemokine (C-X3-C motif) ligand 1
ATP13A4	175	-37.44	ATPase type 13A4
PLAC9	176	-31.53	placenta-specific 9
APOL4	176	-32.77	apolipoprotein L, 4
FLRT2	176	-33.5	fibronectin leucine rich transmembrane protein 2
HNMT	176	-34.85	histamine N-methyltransferase
ZNF395	176	-37.44	zinc finger protein 395
FLT4	177	-37.13	fms-related tyrosine kinase 4
DAB2IP	177	-37.34	DAB2 interacting protein
CCDC38	178	-23.37	coiled-coil domain containing 38
CTNND1	178	-31.5	catenin (cadherin-associated protein), delta 1
POM121	178	-33.73	POM121 transmembrane nucleoporin
ARID5A	178	-34.49	AT rich interactive domain 5A (MRF1-like)
TMCC3	178	-35.65	transmembrane and coiled-coil domain family 3
CPEB1	178	-39.21	cytoplasmic polyadenylation element binding protein 1
ITSN1	179	-33.29	intersectin 1 (SH3 domain protein)
LIMS2	179	-35.88	LIM and senescent cell antigen-like domains 2
COLEC11	179	-36.21	collectin sub-family member 11
FOXO4	179	-40.14	forkhead box O4
KLB	180	-32.29	klotho beta
VPRBP	180	-37.68	Vpr (HIV-1) binding protein
ALS2CL	184	-44.47	ALS2 C-terminal like
SEMA3G	186	-31.39	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3G
H6PD	188	-41.79	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)
List of gene targets for hsa_piR_019914			
CXCL12	170	-27.73	chemokine (C-X-C motif) ligand 12
GRHL1	170	-30.36	grainyhead-like 1 (Drosophila)
PLCD3	170	-34.65	phospholipase C, delta 3

MFI2	170	-35.68	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5
PDE2A	170	-37.29	phosphodiesterase 2A, cGMP-stimulated
TMED4	171	-34.89	transmembrane emp24 protein transport domain containing 4
SEC63	171	-35.44	SEC63 homolog (S. cerevisiae)
XPNPEP3	171	-35.44	X-prolyl aminopeptidase (aminopeptidase P) 3, putative
TXNRD2	171	-36.14	thioredoxin reductase 2
SLC6A14	171	-36.16	solute carrier family 6 (amino acid transporter), member 14
KIF1B	171	-36.73	kinesin family member 1B
ANKDD1A	171	-37.17	ankyrin repeat and death domain containing 1A
GHR	171	-37.39	growth hormone receptor
LEPR	172	-32.43	leptin receptor
ITM2C	172	-38.62	integral membrane protein 2C
RRAD	172	-40.81	Ras-related associated with diabetes
EPB41L4B	172	-42.09	erythrocyte membrane protein band 4.1 like 4B
ZNF506	173	-30.73	zinc finger protein 506
DAB2IP	173	-43.79	DAB2 interacting protein
ATP1B2	174	-38.89	ATPase, Na+/K+ transporting, beta 2 polypeptide
CDC14B	175	-37.63	cell division cycle 14B
CA5B	175	-37.66	carbonic anhydrase VB, mitochondrial
HSD17B13	175	-37.66	hydroxysteroid (17-beta) dehydrogenase 13
IGF1	175	-37.66	insulin-like growth factor 1 (somatomedin C)
IL17RD	175	-37.66	interleukin 17 receptor D
LY75	175	-37.66	lymphocyte antigen 75
LYRM4	175	-37.66	LYR motif containing 4
PER2	175	-37.66	period circadian clock 2
PHLDA1	175	-37.66	pleckstrin homology-like domain, family A, member 1
PYGO1	175	-37.66	pygopus family PHD finger 1
PTCHD1	175	-37.66	patched domain containing 1
PSTPIP2	175	-37.66	proline-serine-threonine phosphatase interacting protein 2
SYNPO2	175	-37.66	synaptopodin 2
EREG	175	-38.02	epiregulin
ALG9	175	-38.15	ALG9, alpha-1,2-mannosyltransferase

ANTXR2	175	-38.23	anthrax toxin receptor 2
ZC3H6	175	-38.58	zinc finger CCCH-type containing 6
ARRB1	176	-36.37	arrestin, beta 1
KCNH8	177	-26.99	potassium voltage-gated channel, subfamily H (eag-related), member 8
ACO1	177	-43.07	aconitase 1, soluble
NARG2	179	-38.72	NMDA receptor regulated 2
PTPN3	179	-39.29	protein tyrosine phosphatase, non-receptor type 3
EHD2	179	-42.58	EH-domain containing 2
B3GAT1	180	-37.43	beta-1,3-glucuronyltransferase 1 (glucuronosyltransferase P)
CLPX	183	-42.5	caseinolytic mitochondrial matrix peptidase chaperone subunit

Table 9.4 Genomic distribution of tRNAs

tRNAs embedded within the intronic regions of protein coding genes			
tRNA ID	**mRNA**	**tRNA ID**	**mRNA ID**
chr2.trna17-GluTTC	ABI2	chr1.trna112-GlnCTG	NBPF8
chr6.trna81-LeuTAA	ADGRB3	chr1.trna21-HisGTG	NBPF8
chr4.trna4-GlnTTG	APBB2	chr1.trna111-HisGTG	NBPF8
chr2.trna14-TyrATA	ARPC2	chr1.trna110-GluTTC	NBPF8
chr18.trna4-LysCTT	ATP5A1	chr1.trna22-GlnCTG	NBPF8
chr18.trna2-GlyTCC	ATP8B1	chr1.trna23-GlnCTG	NBPF8
chr18.trna3-GlyTCC	ATP8B1	chr1.trna109-AsnGTT	NBPF8
chr19.trna9-LysTTT	AXL	chr1.trna25-AsnGTT	NBPF8
chr17.trna17-SupTTA	BCAS3	chr1.trna26-AsnGTT	NBPF8
chr1.trna8-LysCTT	BMP8A	chr1.trna121-GlnCTG	NBPF9
chr7.trna3-ArgCCT	C7orf55	chr12.trna16-GlnTTG	NCKAP5L
chr8.trna11-SerAGA	C8orf37-AS1	chr8.trna3-GluCTC	NSMAF
chr17.trna24-ThrAGT	CEP95	chr14.trna14-GluTTC	NUBPL
chr12.trna7-AspGTC	CLIP1	chr6.trna85-GluTTC	PACRG
chr17.trna36-ThrAGT	CTC1	chr22.trna1-SeC(e)TCA	PARVB
chr17.trna35-SerAGA	CTC1	chr1.trna14-ArgCCT	PDZK1P1
chr17.trna34-IleAAT	CTC1	chr1.trna18-ArgCCT	PDZK1P1
chr13.trna6-GlnTTG	DCLK1	chr6.trna88-PheGAA	PHIP
chr1.trna64-GluTTC	DNM3	chr12.trna1-LysTTT	PPFIBP
chr16.trna22-MetCAT	FBXO31	chr19.trna7-LysTTT	RCN3
chr1.trna127-CysGCA	FNBP1L	chr15.trna6-LysCTT	SCAPER
chr2.trna1-GlyCCC	GREB1	chr20.trna7-GluCTC	SEL1L2
chr17.trna4-ArgTCT	HES7	chr15.trna9-HisGTG	SHF
chr17.trna18-ArgCCT	KCTD2	chr15.trna8-HisGTG	SHF
chr17.trna21-ArgCCT	KCTD2	chr15.trna1-HisGTG	SHF
chr17.trna19-ArgTCG	KCTD2	chr9.trna3-TrpCCA	SNX30
chr12.trna4-AspGTC	LTA4H	chr4.trna5-SupTTA	SORCS2
chr11.trna7-ArgCCT	MARK2	chr3.trna12-CysGCA	TBC1D5
chr8.trna12-TyrGTA	MTFR1	chr12.trna9-IleAAT	TMEM132D
chrX.trna2-LeuTAA	MTRNR2L10	chr1.trna137-Undet???	TNFRSF9
chr20.trna4-LeuTAA	MTRNR2L3	chr17.trna13-GlyCCC	ULK2
chr1.trna13-AsnGTT	NBPF10	chr16.trna25-GlyGCC	VAC14
chr1.trna119-LysCTT	NBPF10	chr16.trna24-GlyGCC	VAC14
chr1.trna118-HisGTG	NBPF10	chr16.trna18-GlyGCC	VAC14

chr1.trna117-GlyTCC	NBPF10	chr16.trna19-GlyGCC	VAC14
chr1.trna116-GluCTC	NBPF10	chr7.trna8-CysGCA	ZNF767P
chr1.trna12-AsnGTT	NBPF20	chr7.trna9-TyrGTA	ZNF767P
chr1.trna122-AsnGTT	NBPF20	chr7.trna10-CysGCA	ZNF767P
chr1.trna108-AsnGTT	NBPF25P	chr7.trna20-CysGCA	ZNF767P
chr1.trna107-AsnGTT	NBPF25P	chr7.trna19-CysGCA	ZNF767P
chr1.trna27-GlnCTG	NBPF25P	chr7.trna11-CysGCA	ZNF767P
chr1.trna15-GlnCTG	NBPF25P	chr7.trna12-SerAGA	ZNF767P
chr1.trna115-AsnGTT	NBPF25P	chr7.trna18-CysGCA	ZNF767P
tRNAs embedded within the intronic regions of lncRNAs			
tRNA	Nearest Feature	tRNA	Nearest Feature
chr1.trna134-GluTTC	ENSG00000233421.3	chr18.trna1-GluCTC	ENSG00000267193.1
chr1.trna5-GluTTC	ENSG00000228549.2	chr18.trna2-GlyTCC	ENSG00000267787.1
chr1.trna9-ArgTCT	ENSG00000260464.1	chr18.trna3-GlyTCC	ENSG00000267787.1
chr1.trna125-AsnGTT	ENSG00000223804.1	chr2.trna4-LeuAAG	ENSG00000233862.1
chr1.trna10-AsnGTT	ENSG00000231734.4	chr2.trna25-GluCTC	ENSG00000236469.1
chr1.trna11-AsnGTT	ENSG00000235398.4	chr3.trna11-ArgACG	ENSG00000230530.1
chr1.trna123-AsnGTT	ENSG00000235398.4	chr4.trna3-CysGCA	ENSG00000251526.1
chr1.trna12-AsnGTT	ENSG00000236943.2	chr5.trna22-AspGTC	ENSG00000231185.2
chr1.trna122-AsnGTT	ENSG00000236943.2	chr5.trna13-ThrTGT	ENSG00000248473.1
chr1.trna120-GlnCTG	ENSG00000233396.3	chr6.trna20-AlaAGC	ENSG00000261584.1
chr1.trna109-AsnGTT	ENSG00000215863.2	chr6.trna29-IleTAT	ENSG00000224843.2
chr1.trna24-AsnGTT	ENSG00000215863.2	chr6.trna64-GlnTTG	ENSG00000246350.1
chr1.trna30-AsnGTT	ENSG00000232151.5	chr6.trna102-AlaAGC	ENSG00000225173.1
chr1.trna95-AsnGTT	ENSG00000232151.5	chr6.trna91-AlaAGC	ENSG00000272541.1
chr1.trna94-GluTTC	ENSG00000234232.2	chr6.trna82-GlyGCC	ENSG00000237494.1
chr1.trna93-AsnGTT	ENSG00000234232.2	chr8.trna13-LeuTAA	ENSG00000253314.1
chr1.trna92-PheGAA	ENSG00000234232.2	chr8.trna4-TyrGTA	ENSG00000253190.2
chr1.trna48-AspGTC	ENSG00000273112.1	chr8.trna5-TyrGTA	ENSG00000253190.2
chr14.trna7-ArgACG	ENSG00000257285.1	chr8.trna6-AlaAGC	ENSG00000253190.2
chr15.trna9-HisGTG	ENSG00000259932.1	chr8.trna11-SerAGA	ENSG00000253773.1
chr15.trna8-HisGTG	ENSG00000260035.1	chrX.trna5-IleGAT	ENSG00000205663.5
chr15.trna4-ArgTCG	ENSG00000261441.1	chr6.trna101-AlaAGC	ENSG00000225595.2
chr16.trna32-LysCTT	ENSG00000261889.1	chr1.trna44-AspGTC	ENSG00000273112.1
chr16.trna29-ProAGG	ENSG00000262521.1	chr1.trna68-GlyGCC	ENSG00000273112.1
chr16.trna9-ProAGG	ENSG00000262668.1	chr1.trna89-AsnGTT	ENSG00000234232.2
chr16.trna10-LysCTT	ENSG00000262668.1	chr1.trna47-AsnGTT	ENSG00000273112.1

chr16.trna11-ProAGG	ENSG00000262668.1	chr1.trna91-GlyCCC	ENSG00000234232.2
chr16.trna12-ArgCCT	ENSG00000262668.1	chr1.trna46-AspGTC	ENSG00000273112.1
chr16.trna13-LysCTT	ENSG00000262668.1	chr1.trna90-ValCAC	ENSG00000234232.2
chr17.trna39-TrpCCA	ENSG00000178977.3	chr1.trna45-GlyTCC	ENSG00000273112.1
chr17.trna10-GlyTCC	ENSG00000178977.3	chr1.trna67-LeuCAG	ENSG00000273112.1
chr17.trna38-AspGTC	ENSG00000178977.3	chrX.trna6-IleGAT	ENSG00000234449.2
chr17.trna37-ProCGG	ENSG00000178977.3	chr1.trna18-ArgCCT	ENSG00000215859.4

Table 9.5 tRNAs harboring regulatory RNAs

tRNAs harboring piRNAs	
tRNA ID	**piRNA ID**
chr1.trna137-Undet???	hsa_piR_004150
chr1.trna7-AsnGTT	hsa_piR_016239
chr1.trna9-ArgTCT	hsa_piR_022650
chr1.trna119-LysCTT	hsa_piR_016742
chr1.trna118-HisGTG	hsa_piR_004987
chr1.trna15-GlnCTG	hsa_piR_004281
chr1.trna98-ValCAC	hsa_piR_016745
chr1.trna32-MetCAT	hsa_piR_016984
chr1.trna68-GlyGCC	hsa_piR_000291
chr11.trna12-ProTGG	hsa_piR_020388
chr11.trna11-LysTTT	hsa_piR_019949
chr12.trna13-AlaTGC	hsa_piR_020485
chr14.trna2-LeuTAG	hsa_piR_020541
chr14.trna19-TyrGTA	hsa_piR_017295
chr14.trna7-ArgACG	hsa_piR_002468
chr14.trna12-LysTTT	hsa_piR_005660
chr14.trna10-IleAAT	hsa_piR_020364
chr15.trna3-CysGCA	hsa_piR_020498
chr17.trna5-GlyGCC	hsa_piR_009295
chr17.trna34-IleAAT	hsa_piR_021214
chr19.trna1-AsnGTT	hsa_piR_009502
chr19.trna8-SeC(e)TCA	hsa_piR_019912
chr2.trna2-TyrGTA	hsa_piR_005019
chr2.trna3-AlaAGC	hsa_piR_020496
chr2.trna19-GlyGCC	hsa_piR_000765
chr3.trna2-ValAAC	hsa_piR_004307
chr4.trna3-CysGCA	hsa_piR_020499
chr5.trna15-ValAAC	hsa_piR_020829
chr5.trna13-ThrTGT	hsa_piR_004993
chr6.trna1-GlnCTG	hsa_piR_020582
chr6.trna5-SerAGA	hsa_piR_015249
chr6.trna13-LysCTT	hsa_piR_000794
chr6.trna14-TyrGTA	hsa_piR_004800
chr6.trna16-TyrGTA	hsa_piR_005018
chr6.trna165-IleAAT	hsa_piR_020363

chr6.trna155-LeuTAA	hsa_piR_012443
chr6.trna149-LysTTT	hsa_piR_019951
chr6.trna131-GlnCTG	hsa_piR_020357
chr6.trna65-AlaAGC	hsa_piR_020500
chr6.trna70-AlaCGC	hsa_piR_020793
chr6.trna114-ArgCCG	hsa_piR_012734
chr6.trna101-AlaAGC	hsa_piR_020497
chr6.trna87-GluCTC	hsa_piR_017716
chr8.trna10-MetCAT	hsa_piR_016980
chrX.trna7-IleGAT	hsa_piR_020362
tRNAs harboring miRNAs	
tRNA ID	**miRNA ID**
chr17.trna7-SerGCT	hsa-miR-4521

Table 9.6 Correlation matrix for tRNAs associated with Overall Survival

tRNAs significant for OS	chr6.trna166-AlaAGC	chr12.trna8-AlaTGC	chr16.trna2-ArgCCT	chr9.trna4-ArgTCT	chr17.trna10-GlyTCC	chr14.trna2-LeuTAG	chr6.trna143-LysTTT
chr6.trna166-AlaAGC	1						
chr12.trna8-AlaTGC	0.56	1					
chr16.trna2-ArgCCT	-0.02	-0.09	1				
chr9.trna4-ArgTCT	0.33	0.27	0.15	1			
chr17.trna10-GlyTCC	0.29	0.41	0.34	0.22	1		
chr14.trna2-LeuTAG	0.49	0.63	0.13	0.39	0.4	1	
chr6.trna143-LysTTT	0.37	0.38	0.19	0.63	0.21	0.41	1
chr6.trna5-SerAGA	0.63	0.66	0.11	0.39	0.5	0.83	0.44
chr6.trna50-SerAGA	0.63	0.68	0.07	0.36	0.51	0.82	0.43
chr6.trna147-SerAGA	0.61	0.65	0.18	0.38	0.52	0.82	0.41
chr6.trna145-SerAGA	0.6	0.61	0.06	0.38	0.52	0.81	0.37
chr6.trna51-SerTGA	0.59	0.63	0.12	0.41	0.5	0.83	0.45
chr6.trna172-SerTGA	0.61	0.63	0.04	0.35	0.48	0.79	0.39
chr6.trna148-SerTGA	0.58	0.68	0.11	0.42	0.49	0.83	0.48

chr6.trna148-SerTGA	chr6.trna172-SerTGA	chr6.trna51-SerTGA	chr6.trna145-SerAGA	chr6.trna147-SerAGA	chr6.trna50-SerAGA	chr6.trna5-SerAGA
						1
					1	0.96
				1	0.94	0.95
			1	0.89	0.93	0.92
		1	0.93	0.94	0.95	0.95
	1	0.91	0.88	0.92	0.92	0.9
1	0.9	0.94	0.9	0.94	0.93	0.94

Boxes highlighted in purple indicate correlation co-efficient ≥ 0.8

Table 9.7 Correlation matrix for tRNAs associated with Recurrence Free Survival

tRNAs significant for RFS	chr6.trna166-AlaAGC	chr1.trna116-GluCTC	chr1.trna59-GluCTC	chr1.trna71-GluCTC	chr1.trna74-GluCTC	chr1.trna77-GluCTC	chr1.trna80-GluCTC	chr6.trna77-GluCTC	chr6.trna87-GluCTC	chr1.trna133-GlyCCC	chr2.trna19-GlyGCC	chr6.trna128-GlyGCC	chr1.trna118-HisGTG	chr6.trna152-ValCAC
chr6.trna166-AlaAGC	1													
chr1.trna116-GluCTC	0.13	1												
chr1.trna59-GluCTC	0.11	0.98	1											
chr1.trna71-GluCTC	0.13	0.97	0.97	1										
chr1.trna74-GluCTC	0.12	0.97	0.97	0.96	1									
chr1.trna77-GluCTC	0.14	0.96	0.97	0.97	0.97	1								
chr1.trna80-GluCTC	0.12	0.97	0.97	0.97	0.97	0.97	1							
chr6.trna77-GluCTC	0.09	0.97	0.98	0.97	0.97	0.97	0.97	1						
chr6.trna87-GluCTC	0.12	0.97	0.97	0.97	0.95	0.95	0.97	0.97	1					
chr1.trna133-GlyCCC	-0.08	0.61	0.62	0.61	0.62	0.64	0.62	0.62	0.62	1				
chr2.trna19-GlyGCC	-0.09	0.62	0.63	0.61	0.62	0.65	0.63	0.63	0.63	0.63	1			
chr6.trna128-GlyGCC	-0.09	0.61	0.62	0.61	0.62	0.64	0.62	0.62	0.62	0.62	0.62	1		
chr1.trna118-HisGTG	0.11	0.67	0.67	0.68	0.62	0.65	0.66	0.66	0.66	0.66	0.66	0.66	1	
chr6.trna152-ValCAC	0.02	0.51	0.51	0.52	0.48	0.5	0.52	0.52	0.52	0.52	0.52	0.52	0.52	1

chr6.trna152-ValCAC	chr1.trna118-HisGTG	chr6.trna128-GlyGCC	chr2.trna19-GlyGCC	chr1.trna133-GlyCCC	chr6.trna87-GluCTC	chr6.trna77-GluCTC
						1
					1	0.95
				1	0.64	0.64
			1	0.99	0.65	0.65
		1	0.99	0.99	0.64	0.64
	1	0.23	0.67	0.24	0.65	0.66
1	0.22	0.51	0.5	0.5	0.47	0.5

Boxes highlighted in purple indicate correlation ≥ 0.8

Table 9.8 List of snoRNAs derived from the intronic regions of protein coding genes

snoRNA ID	Protein coding genes	snoRNA ID	Protein coding genes
ACA59.2-201	CCDC141	SNORA15.2-201	CCT6P3
RNU105B-201	PLCB1	SNORA15.3-201	CLTCL1
SCARNA11-201	CHD4	SNORA15-201	CCT6A
SCARNA1-201	PPP1R8	SNORA16.2-201	N4BP2L1
SCARNA12-201	PHB2	SNORA16B-201	PPP2R5A
SCARNA13-201	SNHG10	SNORA18.6-201	LOC101929645
SCARNA14-201	TIPIN	SNORA19-201	EIF3A
SCARNA15.2-201	XPO5	SNORA2.3-201	C2orf48
SCARNA16.2-201	TEAD1	SNORA20.3-201	POU6F2
SCARNA17-201	SNHG22	SNORA20-201	TCP1
SCARNA18.2-201	SNHG22	SNORA21.1-201	RPL23
SCARNA18-201	TMEM167A	SNORA21-201	RPL23
SCARNA20-201	USP32	SNORA22.1-201	SLC2A13
SCARNA21.2-201	DNAJC16	SNORA22.2-201	CCT6A
SCARNA21-201	CHD3	SNORA22.3-201	CCT6P3
SCARNA22-001	WHSC1	SNORA22-201	CCT6P1
SCARNA23-201	POLA1	SNORA23-201	IPO7
SCARNA3-201	RFWD2	SNORA24.1-201	PARP16
SCARNA4.1-201	ARHGEF4	SNORA25.18-201	CPEB3
SCARNA4-201	KIAA0907	SNORA25.5-201	XPO6
SCARNA5-201	ATG16L1	SNORA25.6-201	ALPK3
SCARNA6-201	ATG16L1	SNORA26.1-201	INTS7
SCARNA7-201	KPNA4	SNORA26.5-201	PCNA
SCARNA8-201	HAUS6	SNORA26-201	DANCR
SNORA1.1-201	LYN	SNORA27-201	RPL21,RPL21P28
SNORA1.2-201	CACNG3	SNORA28-201	EIF5
SNORA1.3-201	NBEAL1	SNORA2A-201	KANSL2
SNORA1.4-201	NAV2	SNORA2B-201	KANSL2
SNORA10-201	RPS2	SNORA30.2-201	FOCAD
SNORA11.2-201	NUDT13	SNORA30-201	SRCAP
SNORA11.3-201	PDE5A	SNORA31.10-201	NLGN4X
SNORA11.4-201	KCNE1	SNORA31.12-201	ALDH3A2
SNORA11.5-201	TRO	SNORA31.13-201	C1orf141
SNORA11-201	MAGED2	SNORA31.16-201	LAMC3
SNORA11B-201	C14orf159	SNORA31.25-201	TPT1
SNORA11C-201	ZNF157	SNORA31.6-201	RPGR
SNORA11D-201	MAGED4,MAGED4B	SNORA31-001	TPT1
SNORA11E-201	MAGED4,MAGED4B	SNORA3-201	RPL27A
SNORA12.2-201	ASAP1	SNORA33-201	RPS12
SNORA12-201	CWF19L1	SNORA34-201	KANSL2
SNORA13-201	EPB41L4A-AS1	SNORA35.3-201	LRCH2
SNORA14.1-201	IL2RA	SNORA35-201	HTR2C
SNORA14A-201	POR	SNORA36.1-201	ASCC1
SNORA14B-201	TOMM20	SNORA36.2-201	GPN1
SNORA15.1-201	CCT6P1	SNORA36A-201	DKC1

SNORA36B-201	RAB3GAP2	SNORA5C-201	TBRG4
SNORA36C-201	AAK1	SNORA60-201	SNHG11
SNORA37-201	MBD2	SNORA62.2-201	ZMYM4
SNORA38-201	PRRC2A	SNORA62.4-201	GNRHR
SNORA38B-201	NOL11	SNORA6-201	RPSA
SNORA40.1-201	RAB3GAP1	SNORA62-201	RPSA
SNORA40.14-201	C5orf17	SNORA63.12-201	INPP5B
SNORA40.8-201	FKBP5	SNORA63.6-201	EIF4A2
SNORA40.9-201	AMD1	SNORA63.8-201	OSCP1
SNORA41.1-201	HMGN2P46	SNORA63-201	EIF4A2
SNORA41-201	EEF1B2	SNORA64.3-201	EIF1B-AS1,MYRIP
SNORA42.3-201	AIMP2	SNORA64-201	RPS2
SNORA4-201	EIF4A2	SNORA65-201	RPL12
SNORA42-201	KIAA0907	SNORA66-201	FAM69A,RPL5
SNORA44.1-201	CGN	SNORA67.1-201	ABL2
SNORA45-201	RPL27A	SNORA67.5-201	ABL2
SNORA46.1-201	AHSA1	SNORA68.1-201	ANKRD27
SNORA46-201	CNOT1	SNORA68-201	RPL18A
SNORA47-201	ZBED3	SNORA69-201	RPL39
SNORA48.12-201	CUX1	SNORA7.3-201	PPME1
SNORA48.6-201	NFATC3	SNORA7.4-201	STIM1
SNORA48.7-201	TMED11P	SNORA70.13-201	LOC101928103
SNORA48.8-201	LOC100128714	SNORA70-201	RPL10
SNORA48-201	EIF4A1,SENP3-EIF4A1	SNORA70B-201	USP34
SNORA49-201	EP400	SNORA70C-201	ASTN2
SNORA5.1-201	MYLK	SNORA70D-001	PHLPP2
SNORA50-201	CNOT1	SNORA70G-201	RAP1B
SNORA51.11-201	DDX60L	SNORA71A-201	SNHG17
SNORA51.1-201	ELMO1	SNORA71C-201	SNHG17
SNORA51.3-201	WDR17	SNORA71D-201	SNHG17
SNORA51.4-201	N4BP2	SNORA72.1-201	ECT2
SNORA51.9-201	FAM69A	SNORA72.3-201	NUCKS1
SNORA51-201	NOP56	SNORA72-201	RPL30
SNORA52-201	RPLP2	SNORA73.1-201	MIB1
SNORA53-201	SLC25A3	SNORA73.3-201	RNGTT
SNORA54-201	NAP1L4	SNORA73B-201	RCC1,SNHG3
SNORA55-201	PABPC4,LOC101929516	SNORA74.1-201	TIMM23,TIMM23B
SNORA56-201	DKC1	SNORA74.2-201	TIMM23B
SNORA57-201	LBHD1	SNORA74.3-201	CDRT1
SNORA58.1-201	NDC1	SNORA74.4-201	MATR3,SNHG4
SNORA58.2-201	UBAP2L	SNORA74.5-201	COX10
SNORA58-001	MRPL3	SNORA74A-201	MATR3,SNHG4
SNORA59A-201	VPS13D	SNORA74B-201	ATP6V0E1
SNORA5A-201	TBRG4	SNORA75.2-201	DDX12P
SNORA5B-201	TBRG4	SNORA75.3-201	LOC642846

SNORA75.4-201	DDX11	SNORD114-21-201	SNHG24
SNORA75.6-201	COG7	SNORD114-22-201	SNHG24
SNORA75-201	NCL	SNORD114-23-201	SNHG24
SNORA76.3-201	LARGE	SNORD114-24-201	SNHG24
SNORA77.2-201	PAQR5	SNORD114-25-201	SNHG24
SNORA77.4-201	RANBP1	SNORD115.1-201	SVIL
SNORA77-201	ATP2B4	SNORD115-23-201	PWAR4
SNORA79.1-201	CCNB1IP1	SNORD116.1-201	USH2A
SNORA79-201	GTF2A1	SNORD117-201	ATP6V1G2-DDX39B,DDX39B
SNORA7A-201	RPL32	SNORD118-201	TMEM107
SNORA7B-201	RPL32P3	SNORD119-201	SNRPB
SNORA8.1-201	USP49	SNORD11B-201	NOP58
SNORA8.4-201	DNAH8	SNORD121A-201	UBAP2
SNORA80-201	URB1	SNORD121B-201	UBAP2
SNORA80B-201	ODC1	SNORD12-201	ZFAS1
SNORA81.1-201	SOD1	SNORD123-201	SNHG18
SNORA81-201	EIF4A2	SNORD124-201	MED24
SNORA84-201	IARS	SNORD125-201	AP1B1
SNORA9.2-201	CSTF2	SNORD126-201	CCNB1IP1
SNORA9.3-201	DDX55	SNORD127-201	PRPF39
SNORD100-201	RPS12	SNORD12B-201	ZFAS1
SNORD101-201	RPS12	SNORD12C-201	ZFAS1
SNORD10-201	EIF4A1,SENP3-EIF4A1	SNORD14.1-201	DNAJC27
SNORD102-201	RPL21,RPL21P28	SNORD14A-201	RPS13
SNORD103A-201	PUM1	SNORD14B-201	RPS13
SNORD103B-201	PUM1	SNORD14C-201	HSPA8
SNORD104-201	SNHG25	SNORD14D-201	HSPA8
SNORD105-201	PPAN,PPAN-P2RY11	SNORD14E-201	HSPA8
SNORD105B-201	PPAN,PPAN-P2RY11	SNORD15A-201	RPS3
SNORD110-201	NOP56	SNORD15B-201	RPS3
SNORD111-201	SF3B3	SNORD16-201	RPL4
SNORD111B-201	SF3B3	SNORD17-201	SNX5
SNORD112.31-201	SYBU	SNORD18.1-201	BLM
SNORD11-201	NOP58	SNORD18A-201	RPL4
SNORD113.2-201	SNHG24	SNORD18B-201	RPL4
SNORD113.3-201	SNHG24	SNORD18C-201	RPL4
SNORD114-11-201	SNHG24	SNORD19.1-201	GNL3
SNORD114-12-201	SNHG24	SNORD19-201	GNL3
SNORD114-13-201	SNHG24	SNORD19B.1-201	GNL3
SNORD114-14-201	SNHG24	SNORD19B-201	GNL3
SNORD114-15-201	SNHG24	SNORD1B-201	SNHG16
SNORD114-16-201	SNHG24	SNORD20-201	NCL
SNORD114-17-201	SNHG24	SNORD21-201	FAM69A,RPL5
SNORD114-19-201	SNHG24	SNORD2-201	EIF4A2
SNORD114-20-201	SNHG24	SNORD23-201	GLTSCR2

SNORD24-201	RPL7A	SNORD58B-201	RPL17
SNORD28.2-201	ALPK2	SNORD58C-201	RPL17,RPL17-C18orf32
SNORD31.1-201	FAM155A	SNORD59A-201	ATP5B
SNORD32A-201	RPL13A,RPL13AP5	SNORD60-201	SNHG19
SNORD33-201	RPL13A,RPL13AP5	SNORD61-201	RBMX
SNORD34-201	RPL13A,RPL13AP5	SNORD62.1-201	MAN1B1
SNORD35A-201	RPL13A,RPL13AP5	SNORD62A-001	PRRC2B
SNORD35B-201	RPS11	SNORD62B-001	PRRC2B
SNORD36A-201	RPL7A	SNORD63.1-201	HSPA9
SNORD36B-201	RPL7A	SNORD63-201	HSPA9
SNORD36C-201	RPL7A	SNORD64.1-201	C1orf204
SNORD37-201	EEF2	SNORD65.3-201	IGF2BP3
SNORD38.2-201	DCP1A	SNORD66-201	EIF4G1
SNORD38.3-201	ADAM9	SNORD67-201	CKAP5
SNORD38A-201	RPS8	SNORD68-201	RPL13
SNORD38B-201	RPS8	SNORD69-201	GNL3
SNORD39.1-201	RAMP1	SNORD70.1-201	NOP58
SNORD41-201	TNPO2	SNORD70-201	NOP58
SNORD42A-201	RPL23A	SNORD71-201	AP1G1
SNORD42B-201	RPL23A	SNORD72-201	RPL37
SNORD43.1-201	CHRDL2	SNORD73.1-201	RPS3A
SNORD43-201	RPL3	SNORD73A-201	RPS3A
SNORD45.1-201	ANKHD1,ANKHD1-EIF4EBP3	SNORD74.1-201	AGBL1
SNORD45.3-201	BTBD9	SNORD74.3-201	BTBD11
SNORD45A-201	RABGGTB	SNORD74.4-201	TBC1D19
SNORD45B-201	RABGGTB	SNORD74.6-201	LOC643339
SNORD45C-201	RABGGTB	SNORD77.3-201	RGS22
SNORD46-201	RPS8	SNORD78.2-201	EXOC6B
SNORD48-201	C6orf48	SNORD78-201	GAS5
SNORD4A-201	RPL23A	SNORD81.2-201	PTN
SNORD4B-201	RPL23A	SNORD8-201	CHD8
SNORD5.1-201	MROH5	SNORD82-201	NCL
SNORD5.2-201	RARB	SNORD83A-201	RPL3
SNORD50.1-201	CCDC158	SNORD83B-201	RPL3
SNORD50.2-201	VPS29	SNORD84-201	ATP6V1G2-DDX39B,DDX39B
SNORD51-201	EEF1B2	SNORD85-201	PUM1
SNORD52-201	C6orf48	SNORD86-201	NOP56
SNORD53_SNORD92.1-201	WDR43	SNORD88A-201	C19orf48
SNORD53-201	WDR43	SNORD88B-201	C19orf48
SNORD54-201	RPS20	SNORD88C-201	C19orf48
SNORD55-201	RPS8	SNORD89-201	RNF149
SNORD56-001	NOP56	SNORD90-201	RC3H2
SNORD57-001	NOP56	SNORD9-201	CHD8
SNORD58.1-201	PRPF39	SNORD92-201	WDR43
SNORD58A-201	RPL17,RPL17-C18orf32	SNORD94-201	PTCD3

SNORD95-201	GNB2L1	snoU13.400-201	KIF9
SNORD96A-201	GNB2L1	snoU13.40-201	LOC643339
SNORD96B-201	AMMECR1	snoU13.410-201	SPATA16
SNORD97-201	EIF4G2	snoU13.411-201	AGBL4
SNORD98-201	CCAR1	snoU13.417-201	RNF114
SNORD99-201	SNHG12	snoU13.423-201	SMG1P1
snosnR60_Z15.2-201	LOC101928782	snoU13.44-201	NCALD
snoU109.3-201	UCHL5	snoU13.45-201	TMEM217
snoU109.4-201	BCAP29	snoU13.454-201	C11orf70
snoU13.101-201	FAM109B	snoU13.473-201	EXD3
snoU13.114-201	CA14	snoU13.476-201	CCDC64
snoU13.117-201	UBR5	snoU13.50-201	ARHGAP42
snoU13.136-201	TLK1	snoU13.57-201	TMC1
snoU13.152-201	PPP1R14C	snoU13.58-201	FOXK2
snoU13.163-201	CCDC25	snoU13.64-201	PRPSAP1
snoU13.17-201	LRRC16A	snoU13.65-201	ZHX1,ZHX1-C8orf76
snoU13.191-201	NUP98	snoU13.66-201	SMCHD1
snoU13.210-201	RALGAPA1,RALGAPA1P	snoU13.82-201	ABCC4
snoU13.219-201	OCLN	snoU13.9-201	DENND4A
snoU13.2-201	LOC100130691	snoU13.93-201	C6orf106
snoU13.222-201	FANCC	snoU2_19.3-201	EIF1AX
snoU13.225-201	SF3B2	snoU2-30.1-201	EIF1AX
snoU13.233-201	LRTOMT	U3.12-201	SKAP1
snoU13.235-201	KIAA2012	U3.13-201	MALRD1
snoU13.240-201	FOXK1	U3.16-201	C20orf194
snoU13.24-201	CCNB1	U3.24-201	LOC101927066
snoU13.245-201	EPRS	U3.27-201	TEX14
snoU13.256-201	SPAG9	U3.31-201	EDA
snoU13.268-201	MOGAT1	U3.4-201	TEX11
snoU13.272-201	BRAT1	U3.44-201	FILIP1
snoU13.279-201	SLC37A3	U3.48-201	FAM83A
snoU13.286-201	XPNPEP3	U3.52-201	RNU5F-1
snoU13.290-201	KDM2A	U3.59-201	GCM1
snoU13.29-201	FMN1	U3.6-201	PTPRM
snoU13.297-201	FGF12	U3.70-201	NUP210L
snoU13.298-201	DNAH12	U3.74-201	PLCD4
snoU13.299-201	WIPF3	U3.75-201	TAF4B
snoU13.303-201	SLC29A3	U3.76-201	METTL5
snoU13.312-201	AKAP12	U3.8-201	CATSPERB
snoU13.330-201	CFAP70	U3.87-201	ETFDH
snoU13.334-201	BCL2	U8.10-201	ULK4P1,ULK4P2
snoU13.354-201	TCF12	U8.11-201	ULK4P1,ULK4P2
snoU13.355-201	ZNF275	U8.17-201	STAB2
snoU13.39-201	PLD1	U8.21-201	AKR1C3
snoU13.394-201	SMG1P2,SMG1P5		

Table 9.9 List of snoRNAs harboring miRNAs

snoRNA ID	Fold change of snoRNAs	mature_miRNA embedded within snoRNAs	Fold change of miRNAs
SNORA36B-201	-1.27	hsa-miR-664a-3p, hsa-miR-664a-5p	-2.93
SNORA81-201	1.27	hsa-miR-1248	-1.08
SNORA84-201	-1.41	hsa-miR-3651	-1.89
SNORA34-201	2.93	hsa-miR-1291	1.9
SNORD125-201	-3.29	hsa-miR-3653	-9.2
SNORA36A-201	1.7	hsa-miR-664b-3p, hsa-miR-664b-5p	1.15

Appendix Table 9.9: Six snoRNAs were found to harbor mature miRNAs. Five snoRNA-miRNA pairs were found to exhibit same direction of expression in tumor tissues, relative to normal tissues (highlighted in green).

Table 9.10 List of snoRNAs harboring piRNAs

Transcript	piRNA_Overlapping Features
SNORD98-201	hsa_piR_000045
SNORD113-8-201	hsa_piR_000330, hsa_piR_019168
SNORD82-201	hsa_piR_000441
SNORD43-201	hsa_piR_000552
SNORD42B-201	hsa_piR_000560
SNORD63-201	hsa_piR_000586
SNORD6-201	hsa_piR_001042
SNORD84-201	hsa_piR_001078
SNORD38B-201	hsa_piR_001101
SNORD51-201	hsa_piR_001152
SNORD113-9-201	hsa_piR_001179
SNORA3-201	hsa_piR_001205
SNORD68-201	hsa_piR_001207
SNORD4A-201	hsa_piR_001346
SNORD71-201	hsa_piR_002158
SNORD83B-201	hsa_piR_002732, hsa_piR_020814
SNORD8-201	hsa_piR_003728
SNORA38-201	hsa_piR_004531
SNORD100-201	hsa_piR_005271, hsa_piR_017104
SNORD67-201	hsa_piR_007635
SNORA5A-201	hsa_piR_010155
SNORD36C-201	hsa_piR_010541
SNORD57-001	hsa_piR_012681
SNORD20-201	hsa_piR_012753
SNORD103A-201	hsa_piR_012925
SNORD103B-201	hsa_piR_012925
SNORD54-201	hsa_piR_013306
SCARNA6-201	hsa_piR_013350
U3.75-201	hsa_piR_016792, region overlaps with 16.901408450704224% of hsa_piR_015150
SNORD114-1-201	hsa_piR_016963, hsa_piR_019201, hsa_piR_019224
SNORD34-201	hsa_piR_016975
SNORD17-201	hsa_piR_017033
SNORD58A-201	hsa_piR_017178
SNORA31-001	hsa_piR_017184
SNORD58B-201	hsa_piR_017194
SNORD85-201	hsa_piR_017458, hsa_piR_020813

SNORD15A-201	hsa_piR_017791
SNORD58C-201	hsa_piR_018292
SNORD1B-201	hsa_piR_018780, hsa_piR_020657
SNORD16-201	hsa_piR_019050
SNORD114-23-201	hsa_piR_019102
SNORD114-3-201	hsa_piR_019166
SNORD113-3-201	hsa_piR_019169
SNORA54-201	hsa_piR_019354
SNORD89-201	hsa_piR_019420
SNORD114-22-201	hsa_piR_019574
SNORD110-201	hsa_piR_019676
SNORD35A-201	hsa_piR_020305
SNORD41-201	hsa_piR_020439
SNORA58-001	hsa_piR_020466
SNORA63.11-201	region overlaps with 10.1010101010101% of hsa_piR_003676
SNORA21-201	region overlaps with 19.548872180451127% of hsa_piR_016664
SNORA28-201	region overlaps with 20.472440944881889% of hsa_piR_018165
SNORA42-201	region overlaps with 21.014492753623188% of hsa_piR_019368
SNORD9-201	region overlaps with 29.523809523809526% of hsa_piR_019911
SNORD62A-001	region overlaps with 33.333333333333329% of hsa_piR_020450
SNORD62B-001	region overlaps with 33.333333333333329% of hsa_piR_020450
SNORD116-14-201	region overlaps with 8.6021505376344098% of hsa_piR_008125

Appendix Table 9.10: Except for piRNAs embedded within SNORA63.11-201, SNORD116-14-201 and U3.75-201, remaining piRNAs were present in the breast tissues. The snoRNA-piRNA pairs (n=35) indicated in yellow are the pairs exhibiting same direction of fold change in tumors, relative to normal tissues.