UNIVERSITÀ DEGLI STUDI DI CATANIA

FACOLTÀ DI MEDICINA E CHIRURGIA
DOTTORATO DI RICERCA IN PATOLOGIA ED EMATOLOGIA
CLINICA, SPERIMENTALE E COMPUTAZIONALE
XXV CICLO

Luciano Cascione

Application of computational and statistical methods to High-throughput gene and microRNA expression to assess their roles in cancer

Tesi di dottorato

Coordinatore:
Chiar.mo Prof. Alfredo Ferro
Relatori:
Dr. Pierluigi Gasparini
Dr. Kay Huebner
Dr. Carlo M. Croce

Anno Accademico 2011/2012

Abstract

High throughput technologies have become a key tool in cancer research. The analysis of gene expression profiles can give insights into changes in proteins pathways that occur during malignant transformation and cancer progression. Transcriptional expression profiling has proven to be a useful and reliable tool for classifying cancers into subgroups that reflect different histopathological characteristics as well as differential prognostic outcome. In the last decades several studies have demonstrated the crucial role of microRNAs (miRNAs) in human disease in particular in cancer. MiRNAs are small non-protein coding RNAs, able to regulate gene expression at post-transcriptional level, binding the 3'UTR of target genes. A single miRNA can regulate the expression of hundreds of target genes, resulting in either theirs degradation or translational repression. The genome-wide profiling of gene expression and microRNAs will allow investigation of genomic changes in cancer development. When mRNA and microRNA levels are measured in the same sample, an integrative analysis can be performed to compare both profiles and determine their interactions. Here I present the integrated analysis of mRNA and miRNAs expression in tumor, adjacent non-tumor (normal) and lymph node metastatic lesion (mets) tissues, from 251 women with Triple Negative Breast Cancers (TNBC). Tissue specific deregulated miRNAs and mRNAs were identified for normal vs tumor vs mets comparisons. We linked specific miRNA signatures to patient overall survival (OS) and distant disease free survival (DDFS). By multivariate analysis the signatures were independent predictors for OS and DDFS. We used miRNA/mRNA anti-correlations to identify clinically and genetically different TNBC subclasses. We also identified miRNA signatures as potential regulators of TNBC subclass-specific gene expression networks defined by expression of canonical signal pathways using IPA Ingenuity software. mRNA expression profiling resulted in clustering of genes expression into 4 molecular subclasses with different expression signatures anti-correlated with the prognostic miRNAs. Our findings suggest that miRNAs have a key role in triple negative breast cancer development probably through their ability to regulate fundamental pathways such as: cellular growth and proliferation, cellular movement and migration. The results also define microRNA expression signatures that characterize and contribute to the phenotypic diversity of TNBC and its metastasis.

To my family and my wife

Acknowledgements

I would like to thank all the people that have helped and supported me during the three-year PhD program; in particular I would like to express my gratitude to my thesis coordinator, Prof. Alfredo Ferro and Dr. Carlo M. Croce for offering me the opportunity to join his lab.

I thank Dr. Alfredo Pulvirenti, for supervising my master and bachelor degree projects. You paved my way to this PhD project.

I would like to thank Rosalba Giugno, Alessandro Lagana e Dario Veneziano for the research we did together.

I would like to thank the members of Carlo Croce's Lab at The Ohio State University for welcoming me in Columbus. I have learned a lot from each one of you; You are great collaborators, it is a real pleasure to work and interact with you.

Many many thanks to Kay Huebner for involving me in the Triple Negative Breast Cancer project, supporting me during my visit in Ohio and always being optimistic about research.

Special thanks to Pierluigi Gasparini for believing in me as bioinformatician, trying to teach me the basic wet-lab techniques, making a tremendous work of editing of my thesis but more importantly for being a good friend.

Finally, I thank my parents and my wife for their love, support and encouragement.

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

Chapter 1

Introduction

1.1 Cancer Biology

In a healthy condition our natural system controls the generation, growth and death (apoptosis) of cells in perfectly balanced equilibrium. In the natural cycle of their life, cells divide to make new tissues as older cells die. Cancer is a heterogeneous group of diseases where this natural system does not work right and cells do not die at the normal rate. This growing mass of cells, the tumor, can skip the entire cell checkpoints and grow uncontrolled. Cancers are generally classified by the type of cells or organ from which they originate. Since malignant growth can occur in virtually all locations of the body, there are over 100 different types of cancers. Cancer is an immensely complex and diverse disease; however, a set of characteristics are shared among almost all malignancies. Those characteristics, named hallmarks of cancer, are a unified set of capabilities that are acquired during tumorigenesis (Figure 1.1). The originally

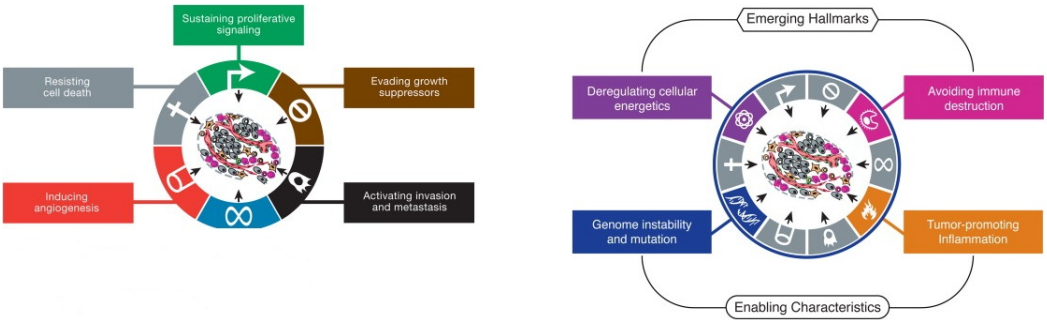

Figure 1.1: The Hallmarks of Cancer. - Left: The original set of hallmarks of cancer). Right: Emerging Hallmarks and Enabling Characteristics. From Hanaham et al. 2000 (1) and Hanaham et al. 2011 (2).

proposed hallmarks of cancer include: growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (2). As a tumor grows, it develops

amazing capabilities to survive, enlarge and spread. It promotes the growth of new blood vessels (angiogenesis) to bring in the oxygen and nutrients it needs. Cancer cells can leave the tumor site and travel through the blood stream and lymphatic system (the network connecting lymph nodes throughout the body) to other parts of the body, such as the liver, lungs or bones. In the new site, cancer cells again may begin to divide and create new tumors. The list of cancer hallmarks has been further extended with newly discovered properties of the tumor such as: deregulating cellular energy balance, decoy immune response, tumor promoting inflammation and genome instability and mutation (2).

1.2 Causes of Cancer

Cancer is often described as the disease of the genome because it acquires its hallmarks through the accumulation of DNA mutations and genome instability (1). Cells can go through uncontrolled growth if there are damages or mutations in their DNA which might compromise the function of the genes involved in cell division. Four types of genes are responsible for the cell division process:

- oncogenes, tell cells when to divide
- tumor suppressor genes, tell cells when not to divide
- suicide genes, control apoptosis and tell the cell to kill itself if something goes wrong
- DNA-repair genes, instruct a cell to repair damaged DNA

Cancer might occurs when mutations allow cells to skip one of these check points, mutations that inhibit oncogene and tumor suppressor gene functions will lead to uncontrolled cell growth. As we age, there is an increase in the number of possible cancer-causing mutations in our DNA. This makes age an important risk factor for cancer. However, it is estimated that only 5-10% of cancer are caused by inherited traits and the remaining 90-95 % are either caused or contributed to by environmental factors (Figure 1.2).

A wide range of substances, Carcinogens, are directly known to be responsible for damaging DNA, promoting or aiding cancer: tobacco, asbestos, arsenic, radiation such as gamma and x-rays. Cancer can also be the result of an inherited genetic predisposition, inherited from family own members. Some mutations are transmittable to the siblings which will make them statistically more likely to develop cancer later in life. Another possible cause of cancer can be virus infections. Worldwide, around 18% of cancers are caused by virus infections such as: human papillomavirus (a cause of cervical cancer), hepatitis B and C (a causes of liver cancer), and Epstein-Barr virus (a cause of some childhood cancers). Human immunodeficiency virus (HIV) - and anything else that suppresses or weakens the immune system, inhibits the body's ability to fight and so increases the chance of developing cancer.

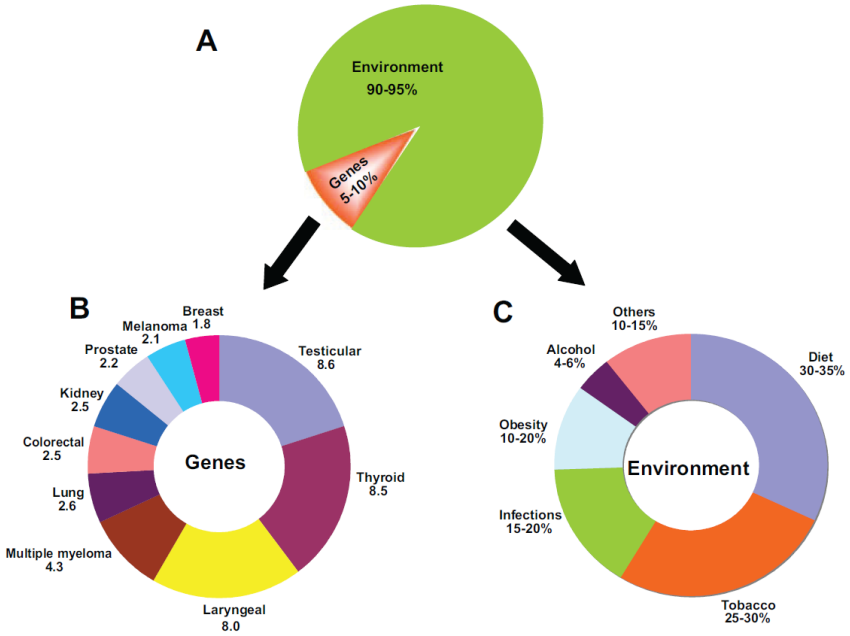

Figure 1.2: The impact of genes and environment on the development of cancer. - A) The percentage contribution of genetic and environmental factors to cancer. B) Numbers represent familial risk ratios - an age-adjusted risk ratio to first-degree relatives of cases compared with the general population. C) Numbers represent the attributable-fraction of cancer deaths due to the specified environmental risk factor. From Anand et al. 2008 (3).

1.3 microRNA in Human Cancer

1.3.1 Bio-genesis and action of microRNAs

MiRNAs are non-coding single-stranded RNAs which are typically 20-25 nucleotides long. As non-coding genes, they are transcribed from DNA, but are not translated into proteins. The mature miRNA molecules are produced in a multi-step process. The DNA sequence is transcribed by RNA polymerase II into a single stranded RNA molecule by hairpin structures known as primary transcripts or pri-miRNAs. The primiRNAs are processed (cutted) into the nucleus by RNAse III Drosha into 70100 nucleotides long fragments called pre-miRNAs.

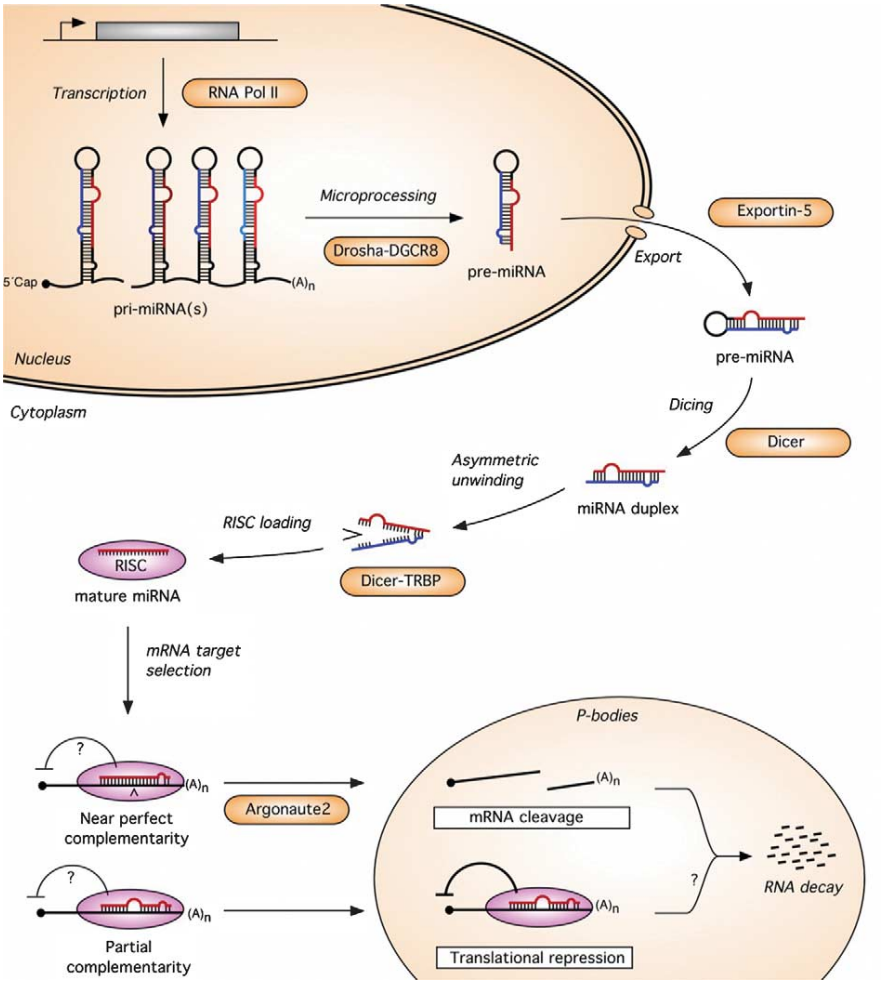

Figure 1.3: A model for the miRNA biogenesis pathway and its action mechanism in mammals. - From Wienholds et al. 2005 (4)

The pre-miRNA molecule is then actively transported to the cytoplasm by a carrier protein. Here, an additional step mediated by the Dicer, generates a double strand

1.3 microRNA in Human Cancer

1.3.1 Bio-genesis and action of microRNAs

MiRNAs are non-coding single-stranded RNAs which are typically 20-25 nucleotides long. As non-coding genes, they are transcribed from DNA, but are not translated into proteins. The mature miRNA molecules are produced in a multi-step process. The DNA sequence is transcribed by RNA polymerase II into a single stranded RNA molecule by hairpin structures known as primary transcripts or pri-miRNAs. The primiRNAs are processed (cutted) into the nucleus by RNAse III Drosha into 70100 nucleotides long fragments called pre-miRNAs.

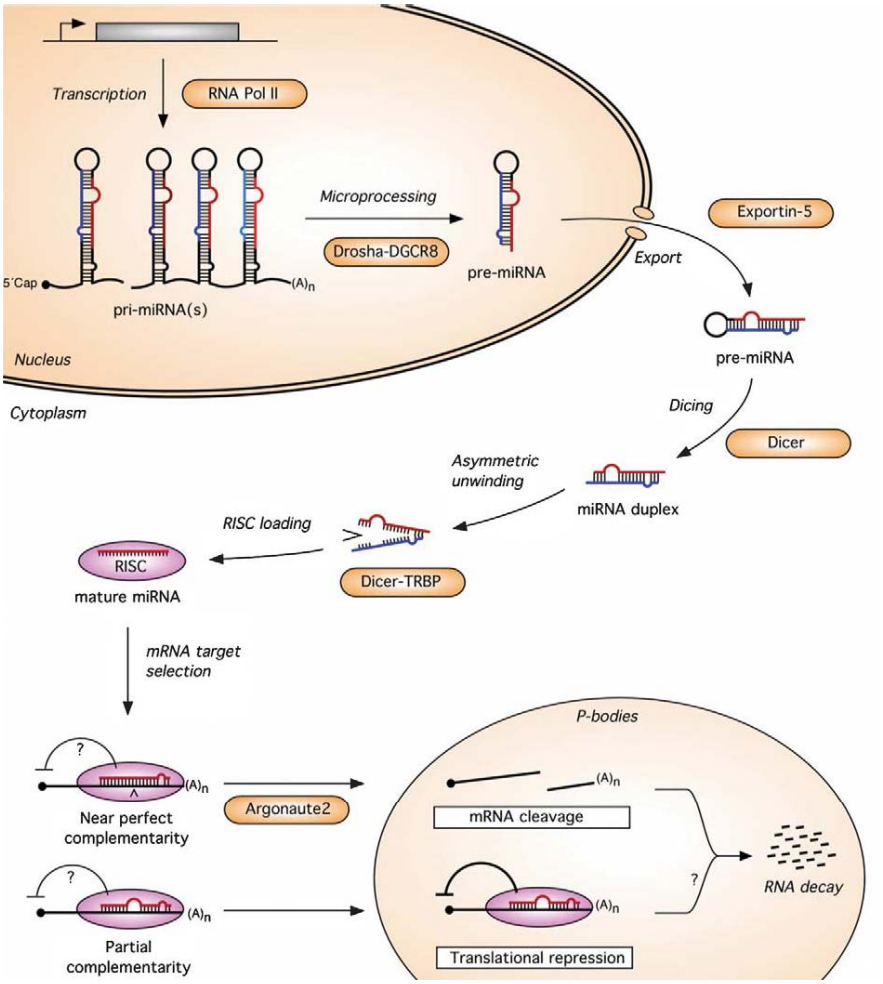

Figure 1.3: A model for the miRNA biogenesis pathway and its action mechanism in mammals. - From Wienholds et al. 2005 (4)

The pre-miRNA molecule is then actively transported to the cytoplasm by a carrier protein. Here, an additional step mediated by the Dicer, generates a double strand

RNA (dsRNA) approximately 22 nucleotides long, including the mature miRNA guide (3p arm) and the complementary passenger strand (5p arm). Once completed the processing steps, through this mechanism that is not fully characterized yet, the mature miRNA is able to regulate gene expression at post-transcriptional level, binding through partial complementarity the 3' untranslated region (3' UTR) of the target mRNA, and mainly leading to either mRNA degradation or translation inhibition (5). Depending on the targeted mRNAs, the miRNA action ultimately results in reduced protein levels and profound consequences on cellular homeostasis. Recent bioinformatics and experimental reports suggest that over 30% of human genes are direct targets of miRNAs (5), indicating their key roles in almost every biological process, including: cell cycle regulation, cell growth, apoptosis, cell differentiation and stress response. Recent genome-wide analyses have identified dysregulated miRNA expression in human malignancies (5), giving them an oncogenic (Table 1.2) or tumor suppressor role (Table 1.1) (6).

Table 1.1: Representative examples of Tumor-suppressor miRNAs in the most common human cancers. From Di Leva et al. 2012 (6).

miRNA	Targets	Tumor	Impact on metastasis	Description
miR-15/16	BCL2	CLL		BCL2 repression by these microRNAs induces apoptosis in a leukemic cell line model
	COX-2	Colon cancer		miR-16 as a central post-transcriptional regulator of COX-2 and shows the ability of elevated levels of HuR to antagonize miR-16 function
	CHEK1	Follicular lymphoma		Distinct microRNA profiles are associated with an increased proliferative capacity and a late germinal center B-cell phenotype
	CEBP, CDC25a, CCNE1	Fibroblast		Upon cell cycle re-entry, the rapid decay of miR-16 alleviates repression of target genes, allowing proper resumption of the cell cycle
	VEGF, VEGFR2, FGFR1	Fibroblast		miR-16 plays important roles in regulating cell-intrinsic angiogenic activity of endothelial cells
	FGF2, FGFR1	Cancer-associated fibroblast		Down-regulation of miR-15 and miR-16 in cancer-associated fibroblasts (CAFs) promotes tumor growth and progression

Continued on next page

Table 1.1 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	CCNE1			miR-15 and miR-16 families as novel transcriptional targets of E2F, which, in turn, modulates E2F activity
	FGFR1, PI3KCa, MDM4, VEGFa	Multiple myeloma		Deletion of miR-15/16 is commonly observed in early stages of multiple myeloma
	WIP1			Role of miR-16 in the regulation of Wip1 phosphatase in the DNA damage response and mammary tumorigenesis
	BMI-1	Ovarian cancer		Bmi-1 is down-regulated by miR-15a or miR-16 expression and leads to reduction in ovarian cancer cell proliferation and clonal growth
	CCND1, CCND2, CCNE1	Lung cancer		Overexpression of miR-15/16 induces arrest in G(1)-G(0)
miR-31	ITGA5, RDX, RhoA, FZD3, M-RIP, MMP16	Breast cancer	Suppresses	miR-31 uses multiple mechanisms to oppose metastasis
	SATB2	Cancer-associated fibroblast		New insights into tumorstroma interaction and involvement of miR-31 in regulation of tumor cell motility
miR-34	SIRT1	Colon cancer		miR-34 suppression of SIRT1 leads to apoptosis only in colon cancer cells with wild-type p53
	BCL2, NOTCH, HMGA2			miR-34-mediated suppression of self-renewal is related to the direct modulation of the downstream targets Bcl-2, Notch, and HMGA2
	MYC	Fibroblast		During senescence, miR-34a targets the proto-oncogene MYC and co-ordinately controls a set of cell cycle regulators
	AXL	Lung cancer		Axl receptor is regulated by miR-34a and miR-199a/b, suppressed by promoter methylation in solid cancer

Continued on next page

Table 1.1 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	MET	Ovarian cancer		MET is a critical effector of p53, and inhibition of MET may be an effective antimetastatic approach to treat cancers with p53 mutations
	NANOG, SOX2, MYCN	Embryonic fibroblast		Suppression of reprogramming by miR-34a due to repression of pluripotency genes
	SNAIL	Colon cancer		A new link between p53, miR-34, and Snail1 in the regulation of cancer cell EMT programs
miR-143/145	KRAS, RREB1	Pancreatic cancer		miR-143/miR-145 are suppressed by KRAS through RREB1, revealing a feed-forward mechanism that potentiates Ras signaling
	KRAS, MYC, CCND2, CDK6, E2F3	Colon cancer		EGFR suppresses miR-143 and miR-145 in murine models of colon cancer
	BCL2	Cervical cancer		Promotion of apoptosis by miR-143 through the suppression of BCL2
	PAI1	Bladder cancer		miR-145 and PAI1 as clinically relevant biomarkers in bladder cancer
	PRC1, PLK1	Liposarcoma		The down-regulation of PRC1 and its docking partner PLK1 suggests that miR-143 inhibits cytokinesis in these cells
	MLL-AF4	ALL		Therapeutic promise of up-regulating miR-143 expression for MLL-AF4 B-cell ALL
	MMP-13	Osteosarcoma		Down-regulation of miR-143 correlates with the lung metastasis of human osteosarcoma cells by promoting cellular invasion, probably via MMP-13 up-regulation
	ERK5	Burkitt lymphoma		miRs-143 and -145 may be useful as biomarkers that differentiate B-cell malignant cells from normal cells
Let-7 family	KRAS	Lung cancer		The let-7 family negatively regulates let-60/RAS in C. elegans and lung tumors

Continued on next page

Table 1.1 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	HMGA2			Chromosomal translocations associated with human tumors disrupt repression of high mobility group A2 (Hmga2) by let-7 miRNA
	MYC	Burkitt lymphoma		Dysregulation of let-7 participates in genesis and maintenance of Burkitt lymphoma and other MYC-dysregulated cancers
	IMP-1			Let-7-oncofetal proteins could be novel therapeutic targets and potential biomarkers for cancer treatment
	DICER			Existence of a regulatory loop to regulate the equilibrated state of Dicer and various miRNAs
	CDC-34	Fibroblast		Let-7 represses Cdc34, stabilizes Wee1 kinase, and increases a fraction of cells in G(2)/M in primary fibroblasts
	IL6	Breast cancer		Inflammation activates a positive feedback loop that maintains the epigenetic transformed state
	E2F2, CCND2	Prostate cancer		Let-7a acts as a tumor suppressor in prostate cancer by down-regulating E2F2 and CCND2
	BCL-XL	Liver cancer		Let-7 suppresses Bcl-xL expression in hepatocellular carcinomas and potentiates sorafenib-induced apoptosis
	PLC1	Breast cancer		Tumor-suppressor function by negatively regulating EGF-driven cell invasion, viability, and cell cycle progression in breast cancer
miR-200 family	ZEB1, ZEB2	Breast cancer	Suppresses	Down-regulation of the miR-200 family may be an important step in tumor progression
	ERRFI-1	Bladder cancer		miR-200 is sufficient to restore EGFR dependency at least in some of the mesenchymal bladder cancer cells
	ZEB1, CTNNB1	Nasopharyngeal carcinoma		The inhibitory effects of miR-200a on cell growth, migration, and invasion are mediated by distinct targets and pathways

Continued on next page

Table 1.1 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	BMI-1	Pancreatic cancer		ZEB1 links EMT and stemness maintenance by suppressing the miR-200 family and thereby promotes migration
	PLC1	Breast cancer		Tumor-suppressor function by negatively regulating EGF-driven cell invasion, viability, and cell cycle progression in breast cancer
	FAP1			miR-200c sensitizes cells to apoptosis mediated by CD95
	SUZ12	Breast cancer		The miR-200b-Suz12-cadherin pathway is important for cancer stem cell growth and invasive ability
	FLT1	Lung cancer		miR-200 suppresses lung adenocarcinoma metastasis by targeting Flt1 in tumor cells
	JAG1, MALM2, MALM3			These findings explain increased Notch signaling in some types of cancers, where mutations in Notch pathway genes are rare
	FN1, LEPR, NTRK2, ARHGAP19	Breast and endometrial cancer		miR-200c actively represses a program of mesenchymal and neuronal genes involved in cell motility and anoikis resistance
	p38	Ovarian cancer		miR-200a-dependent stress signature correlates with improved survival of patients in response to treatment

Table 1.2: Representative examples of OncomiRs in the most common human cancers. From Di Leva et al. 2012 (6).

miRNA	Targets	Tumor	Impact on metastasis	Description
miR-106a 363, miR-106b 25	BIM, p21	Gastric cancer		The miR-106b-25 cluster is involved in E2F1 post-transcriptional regulation and may play a key role in the development of TGF resistance in gastric cancer
	E2F1	Prostate cancer		microRNA expression becomes altered with the development and progression of prostate cancer. Some of these microRNAs regulate the expression of cancer-related genes in prostate cancer cells
	PTEN	Prostate cancer		Proto-oncogenic miRNA-dependent network for PTEN regulation
miR-21	PTEN	Cholangiocar.	Promotes	miR-21 modulates gemcitabine-induced apoptosis by phosphatase and the tensin homolog deleted on chromosome 10 (PTEN)-dependent activation of PI3-kinase signaling
	TPM1	Breast cancer		Suppression of miR-21 can inhibit tumor growth
	PDCD4	Breast cancer		The tumor suppressor protein programmed cell death 4 (PDCD4) is regulated by miR-21, and it has been demonstrated that PDCD4 is a functionally important target for miR-21 in breast cancer cells
	SPRY1			miR-21-null mice show a significant reduction in papilloma formation compared with wild-type mice due to the up-regulation of its tumor-suppressor targets
	RECK, TIMP3	Glioblastoma		The inhibition of miR-21 provides a novel therapeutic approach for physiological modulation of multiple proteins whose expression is deregulated in cancer

Continued on next page

Table 1.2 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	p63, JMY, TOPORS, TP53BP2, DAXX, HNRPK, TGFRII	Glioblastoma		miR-21 targets multiple important components of p53, transforming growth factor- (TGF), and mitochondrial apoptosis tumor-suppressive pathways
	MARKS	Prostate cancer		miR-21 could promote apoptosis resistance, motility, and invasion in prostate cancer cells
	ANP32A, SACA4	Prostate cancer		
miR-10a/10b	HOXB1, HOXB3	Pancreatic cancer	Promotes	miR-10a is a key mediator of metastatic behavior in pancreatic cancer that regulates metastasis via suppression of HOXB1 and HOXB3
	HOXD10	Breast cancer		TWIST transcription factor induces expression of a specific microRNA that suppresses its direct target and in turn activates another pro-metastatic gene, leading to tumor cell invasion and metastasis
	KLF4	Esophageal cancer		A significant correlation of miR-10b level with cell motility and invasiveness
	TIAM1	Breast cancer		A mechanism for the regulation of Tiam1-mediated Rac activation in breast cancer cells
	Nf1	Ewing's sarcoma		miR-10b may play an important role in NF1 tumorigenesis through targeting neurofibromin and RAS signaling
miR-107/103	DICER	Breast cancer	Promotes	Dicer inhibition drifts epithelial cancer toward a less-differentiated, mesenchymal fate to foster metastasis
miR-9	PRDM1	Lymphomas	Promotes	miRNA-mediated down-regulation of PRDM1/Blimp-1 may contribute to the phenotype maintenance and pathogenesis of lymphoma cells by interfering with normal B-cell terminal differentiation
	CDH1	Breast cancer		

Continued on next page

Table 1.2 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	CAMTA	Glioblastoma		miR-9 is highly expressed in glioblastoma cancer stem cells and reduces the levels of CAMTA tumor-suppressor
miR-1792	TSP-1, CTGF	Colon	Promotes	Up-regulated in colonocytes coexpressing K-Ras, c-Myc and p53 impaired activity
	E2F2, E2F3	Prostate/Burkitt lymphoma/testis carcinoma/		Presence of an autoregulatory feedback loop between E2F factors and miR-17/92
	BIM, PTEN	c-Myc-induced lymphoma		Transgenic mice with higher expression of miR-17/92 in lymphocytes
	HIF1	Lung cancer		Intricate and finely tuned circuit involving c-myc, miR-17/92, and HIF1
	PTPRO	Cervix tumor cell line		PTPRO gene is co-regulated by both E2F1 and miR-17/92 at transcriptional and post-transcriptional level, respectively
	p63	Myeloid cells		miR-92 increases cell proliferation by negative regulation of an isoform of the cell cycle regulator p63
	BIM, PTEN, PRKAA1, PPP2R5e	T-cell acute lymphoblastic leukemia		Functional genomics approach reveals a co-ordinate clamp-down on several regulators of phosphatidylinositol-3-OH kinase-related survival signals by the leukemogenic miR-19
	JAK1	Endothelial cells		The miR-17/92 family may provide an interesting therapeutic perspective specifically to enhance therapeutic angiogenesis
	HBP1	Breast cancer		The miR-17/92 cluster plays an important role in breast cancer cell invasion and migration by suppressing HBP1 and subsequently activating Wnt/-catenin
	p21(WAF1)	Ras-induced senescent fibroblasts		Disruption of senescence by miR-17/92 or its miR-17/20a components leads to enhanced oncogenic transformation by activated Ras in primary human cells

Continued on next page

Table 1.2 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	TGFII SA4	Glioblastoma		miR-17/92 attenuates the TGF signaling pathway to shut down clusterin expression, thereby stimulating angiogenesis and tumor cell growth
	MnSOD, GPX2, TRXR2	Prostate		miR-17/92 may suppress tumorigenicity of prostate cancer through inhibition of mitochondrial antioxidant enzymes
miR-221/222	p27kip1	Glioblastoma, prostate and thyroid carcinoma	Promotes	Certain cancer cell lines require high activity of miR-221/222 to maintain low p27kip1 levels and continuous proliferation
	p57kip2	Normal fibroblast		Up-regulation of miR-221/222 is tightly linked to the initiation of S phase with growth factor signaling pathways that stimulate cell proliferation
	PTEN, TIMP3	Non-small cell lung cancer and hepatocellular carcinoma		miR-221/222, by targeting PTEN and TIMP3 tumor suppressors, induce TRAIL resistance and enhance cellular migration. The MET oncogene is involved in miR-221/222 activation through the c-Jun transcription factor
	FOXO3A	Breast cancer		The miR-221/222 cluster targets FOXO3A to suppress p27kip1 also at a transcriptional level
	KIT	Endothelial cells		Interaction between miR-222 and c-Kit is likely to be part of a complex circuit that controls the ability of endothelial cells to form new capillaries
	ESR1	Breast cancer		Modulation of ER is associated with antiestrogen therapy
	PUMA	Glioblastoma		miR-221/222 directly regulate apoptosis by targeting PUMA in glioblastoma
	TRSP1	Breast cancer		miR-221/222 promote EMT and contribute to the more aggressive clinical behavior of basal-like breast cancers
	PTP	Glioblastoma		miR-221/222 regulate glioma tumorigenesis at least in part through the control of PTP protein expression
	DICER	Breast cancer		Dicer is low in ER-negative breast cancers, since such cells express high miR-221/222

Continued on next page

Table 1.2 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	APAF1	Non-small cell lung cancer		miR-221/222 are modulated by both epidermal growth factor (EGF) and MET receptors, and, by targeting APAF1, miR-221/222 are responsible for gefitinib resistance
miR-155	SOCS1	Breast cancer		miR-155 is an oncomiR in breast cancer, and it has been suggested that miR-155 may serve as a bridge between inflammation and cancer
	CEBPB, PU.1, CUTL1, PICALM	AML		miR-155 as a contributor to physiological GM expansion during inflammation and to certain pathological features associated with AML
	BACH1, ZIC3			The induction of miR-155 by EBV contributes to EBV-mediated signaling in part through the modulation of transcriptional regulatory factors
	ETS1, MEIS1	Human cord blood CD34+		miR-155 is required for megakaryocytic proliferation and differentiation
	C-MAF	Lymphocytes		bic/microRNA-155 plays a key role in the homeostasis and function of the immune system
	HGAL	Diffuse large B-cell lymphoma		Cell dissemination and aggressiveness is a phenotype of DLBCL typically expressing high levels of miR-155 and lacking HGAL expression
	JMJD1A	Nasopharyngeal carcinoma		Up-regulation of miR-155 is partly driven by LMP1 and LMP2A, and results in down-regulation of JMJD1A, associated with N stage and poor prognosis
	WEE1	Breast cancer		miR-155 enhances mutation rates by decreasing the efficiency of DNA safeguard mechanisms by targeting of cell cycle regulators such as WEE1
	TP53INP1	Pancreatic cancer		TP53INP1 expression is repressed by the oncogenic micro RNA miR-155, which is overexpressed in pancreatic carcinoma cells

Continued on next page

Table 1.2 – *Continued from previous page*

miRNA	Targets	Tumor	Impact on metastasis	Description
	SMAD1, SA5, HIVEP2, CEBPB, RUNX2, MYO10			Role for miR-155 in controlling BMP-mediated cellular processes
	FOXO3a	Breast cancer		Molecular links between miR-155 and FOXO3a affect cell survival and response to chemotherapy in breast cancer
	hMSH2, hMSH6, and hMLH1	Colon cancer		Inactivation of mismatch repair is induced by miR-155
	SMAD5	Diffuse large B-cell lymphoma		Highlighted a hitherto unappreciated role of SA5 in lymphoma biology and defined a unique mechanism used by cancer cells to escape TGF's growth-inhibitory effects

Alterations in the expression of miRNAs were initially identified in B-cell leukemia (7), now are considered a common characteristic of all human tumors. Genome-wide profiling showed that miRNA expression signatures (miRNome) allowed different types of cancer to be discriminated with high accuracy (8)(9). Iorio et al. in 2005 (10) compared normal breast tissue with breast cancer tissue using microRNA profile. The overall miRNA expression could clearly separate normal versus cancer tissues; they also could identify miRNAs whose expression was correlated with specific breast cancer bio-pathologic features, such as estrogen and progesterone receptor expression, tumor stage, vascular invasion, or proliferation index.

1.3.2 Genetic abnormalities and miRNAs

Having addressed the strong association between miRNA levels and human diseases, the precise control of miRNAs levels is essential in maintaining normal cellular homeostasis. Genomic abnormalities, such as chromosomal rearrangements, genomic amplifications, deletions or mutations, can alter miRNA genes too reflecting their affects on protein coding genes down stream. In 2004, an in silico study showed that more than half of miRNAs map to genomic regions that are frequently altered in cancer (11):

- loss of heterozygosity regions (LOH) (e.g. miR-15a/16-1)
- amplified regions (e.g. miR-17-92 cluster, miR-155)

- breakpoint regions and fragile sites (FRA) (e.g. let-7 family members)

Amplification of miRNAs might also occurs in cancer, this is exemplified by the human oncogenic cluster miR-17-92, which is located at chromosome 13q31 (12). Overexpression of miR-17-92 increases MYC-induced lymphomagenesis, and this region is preferentially amplified in cancers such as DLBCL, follicular lymphoma, mantle cell lymphoma and primary cutaneous B-cell lymphoma.

1.3.3 Epigenetics and miRNAs

It is now recognized that cancer is mainly a genetic disease, however, genetic lesions alone cannot explain the complexity of the aberrations that arise in cancer cells. Epigenetic, defined as heritable change in gene activity that is independent of DNA sequence, play a prominent role in the initiation and progression of cancer. Three main epigenetic events regulate tumor-associated genes:

- aberrant hypermethylation of tumor-suppressor genes
- global DNA hypomethylation item post-translational modifications of histones

MiRNAs can be target of epigenetic events that, in some instances, can explain the perturbation of miRNA expression in cancer (13). An extensive analysis of genomic sequences of miRNA genes have shown that approximately half of them are associated with CpG islands, suggesting that they could be subjected to this mechanism of regulation (Weber et al, 2007). The miRNAs might not only be regulated by epigentic mechanisms, but miRNAs can also regulate enzymes that are involved in the methylation of the CpG islands in tumor suppressor genes. In conclusion, epigenetic changes complemented by genetic inactivation due to mutation or deletion can shed light on the mechanisms that partially account for the miRNA dysregulation in cancer.

Chapter 2

Data analysis

2.1 High-throughput profiling in Cancers

The development of powerful and scalable methods to analyze nucleic acids has transformed biological inquiry and has the potential to alter the practice of medicine (14),(15). The application of such technologies, together with powerful computational methods in human disease and animal models has facilitated the study of both normal and disease-affected tissues in a manner previously not possible. The cellular and molecular heterogeneity of cancer and the large number of genes involved in controlling cell growth, death, and differentiation emphasize the importance of studying multiple genetic alterations in concert. Gene expression profiling allows the simultaneous measurement of the activity (expression) of thousands of genes in a cancer cell. Molecular profiling is an emerging concept in clinical decision making that involves classification of biological specimens such as tumors or other tissues into groups based on multiple changes at the genomic and transcriptomic levels. In the last decade, molecular profiling technologies have advanced our knowledge of cancer biology. Early cancer genome analysis has already led to the discovery of new targets for cancer therapy and new insights about specific genetic mutations and clinical response, as well as new approaches useful for diagnosis and prognosis. These initial efforts have motivated large-scale coordinated cancer genomic efforts to obtain complete catalogs of the genomic alterations in specific cancer types (The Cancer Genome Atlas [TCGA, http://cancergenome.nih.gov). Microarray technology enables simultaneous measurement of thousands of messenger RNAs transcripts (mRNA). Since all proteins in the cells are produced by the translation of mRNA, the mRNA expression levels provide a good approximation of the abundance of proteins (Figure 2.1).

.

A variety of different approaches are being used to profile the mRNA/microRNA expression levels. They generally involve the amplification of DNA templates by PCR and the physical binding of template DNA to a solid surface or to tiny beads called microbeads. These techniques are often referred to as massively parallel DNA sequencing, because thousands or millions of sequencing reactions are run at once to greatly speed

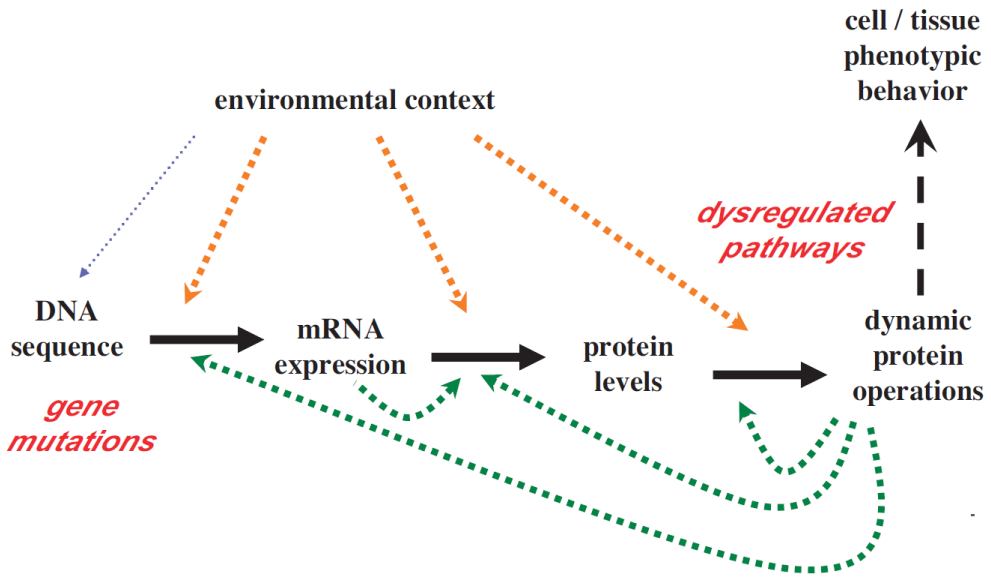

Figure 2.1: The transfer of genetic information from DNA, through mRNA to proteins. - The genetic mutations are translated into dysregulation of cellular pathways, which in turn can impact transcription and/or translation. The biological processes are influenced by environmental context. From Kreeger et al. 2010 (16)

up the process. All next generation sequencing systems use clonal cluster sequencing. The process, which begins with a single target molecule, involves creation of a clonal target during an intermediate amplification step. Multiple identical copies are required to produce a high signal-to-noise-ratio. Next generation sequencing (NGS) technologies provide a digital expression profiling readout that is fundamentally different than analog measurement systems like microarrays. The NanoString company (Seattle, WA, USA) recently introduced a new concept of digital platform; it is not biased by enzymatic steps such as cDNA synthesis and amplification by PCR. The Nanostring technology can be used to detect any type of nucleic acid in solution and could be modified with appropriate recognition probes to detect other biological molecules as well. It is based on direct digital measurement of gene expression through target-specific color coded probes with high level precision and sensitivity at less than one transcript copy per cell (17). Choosing the best platform for mRNA/microRNA profile is always difficult; and it should best based on experience, experimental conditions in laboratory, and more important on goals of research.

2.2 Normalization of expression profiles

All of the most popular and widely used profiling methods face significant introduction of bias due to differences in sample RNA preparation, dye labelling, hybridization and washing efficiency, peculiarities of print tip, spatial or hybridization specific effects

or pre-amplification of extracted RNA. For these reasons normalization is an essential aspect of data processing. It can minimize systematic technical or experimental variation. This variation has significant impact on the detection of differentially expressed molecules between two or more conditions. Inappropriate normalization of the data can lead to incorrect conclusions. Rigorous normalization of miRNA data may even be more critical than that of other RNA functional classes since relatively small changes in miRNA expression may be biologically and clinically significant (18),(19). There is no consensus normalization method for the three miRNA profiling approaches cited above. Several normalization techniques are currently in use, of which some are similar to mRNA profiling normalization methods, while others are specifically modified or developed for miRNA data. The characteristic nature of miRNA molecules, their composition and the resulting data distribution of profiling experiments challenges the selection of adequate normalization techniques. Several studies pointed out that selection of the data preprocessing method can have great impact on the resulting data outcome (20)(21)(22)(23),(24). Thus, prior to normalization, data pre-processing step could be useful. It includes platform and vendor specific steps, such as e.g., baseline adjustment and threshold setting for RT-qPCR analyses, background correction for microarray technology, or filtering for small RNA-sequence data. Following these very first steps of raw data preprocessing, the researcher has to choose the optimal normalization strategy to correct for systematic and technical variation enabling a better estimation of the biological variation.

2.2.1 Normalization approaches for RT-PCR

RT-PCR is generally accepted as gold standard for microRNA measurement and normalized microRNA RT-PCR profiling data is used for evaluation of the goodness of miRNA microarray normalization methods (25)(21). The signal intensities may depend on reverse transcription and PCR reaction efficiencies, thus normalization of profiling data is needed for reflecting true miRNA levels. The common normalization methods for microRNA RT-PCR profiling are based on

- predefined invariant endogenous controls
- reference miRNAs
- small nuclear and small nucleolar RNA

.

Vandesompele et al. (26) argued that it is best to normalize molecules with reference molecules belonging to the same RNA class because the use of small non-coding RNAs other than miRNAs does not mirror the physicochemical properties of miRNA molecules. Using non-miRNA reference genes for qPCR normalization is not suitable when the overall abundance of miRNA varies, e.g., in experiments affecting the miRNA processing machinery or in comparisons involving multiple tissues or combinations of tissues and cell lines (27). Selection of invariant miRNAs identified by algorithms

specifically developed for reference gene evaluation and selection was superior over small non-coding RNA based normalization (28)(18). These algorithms are based on reference gene ranking and stepwise elimination of the least stable gene (26) or repeated pair wise correlation and regression analysis (29), or statistical linear mixed-effects modelling (30) of the respective experimental data. Moreover invariant microRNAs can be selected based on a distinguishable low standard deviation and high-mean population as suggested for miRNA microarray preprocessing for RT-qPCR profiling experiments as well (22). Basically, the use of more than one reference gene increases the accuracy of quantification compared to the use of a single reference gene(26)(30). A new interesting and debated method is the calculation of a plate normalizing factor for RT-qPCR based expression profiling platforms; A scaling method has been suggested by Wang in 2009, which uses the average of eight selected miRNA expression values from a descending sorted list (31). Plate-normalizing factor corresponds to an enlargement of percentile normalization and needs further validation by independent datasets prior to judge the robustness of this method. For large scale microRNA expression profiling studies the mean expression value normalization outperformed the current normalization strategy that makes use of stable small RNA controls, such as e.g., snoRNAs proposed by manufacturers, in terms of better reduction of technical variation (28). However, the selection of a limited number of miRNAs or small RNA controls that resemble the mean expression value can be successfully used for normalization in follow-up studies where only a limited number of miRNA molecules are profiled to allow a more accurate assessment of relevant biological variation from a miRNA RT-qPCR profiling experiment (28)(19).

2.2.2 Normalization methods for microarray experiments

Different normalization methods have been used on miRNA microarray expression profiling data sets, but there is currently no clear consensus about their relative performances [(22). Some have even chosen to omit this key step [(32),(33),(34) but comparative studies on the relative performance of different methods within a miRNA microarray platform have emphasized the need for evaluating and identifying appropriate normalization algorithms (35)(22)(21). Indeed signal intensities of miRNA microarray experiments may be biased by differences in sample RNA preparation, dye labelling, hybridization and washing efficiency, peculiarities of print tip, spatial or hybridization specific effects or pre-amplification of extracted RNA. miRNA microarrays can be single-color or dual-color systems calling for different normalization approaches. Single-color miRNA microarrays have been predominately used, while dual-colour hybridization systems are less frequently prevalent (35). Both can be observed with respect to intra-array normalization for the correction of dye effects and inter-array approaches for the balance of the distribution differences among experiments (36). The first normalization methods to be used with miRNA array data employed centering to median values (37) or scaling based on total array intensities (38). Certain methodologies currently used for large-scale genome arrays have been adapted to and modified for miRNA arrays such as Quantile (39) and LOESS (Locally Weighted Regression and Smooting Scatterplots)

reviewed in (40). Various assumptions are often taken by normalization methods. Scaling, LOESS and Quantile are based on two assumptions,

- only a small portion of spots is differentially expressed,
- differentially expressed spots are homogeneously distributed with respect to both, over- and under-expressed miRNAs (23).

These assumptions could fail for miRNA platforms as they are printed with a relatively small number of selected sequences (23)(21). There are relatively few known microRNAs for any species (approximately 1000 for humans), and the proportion of microRNAs expressed in a given sample tends to be much smaller than for mRNAs (reflected in the tissue-specific expression pattern of many microRNAs) (24). Because of this, the proportion of miRNAs that are differentially expressed (among those expressed at all) is much larger than that observed when profiling global mRNA expression (24). Thus, it needs to verify whether these assumptions hold true for the respective datasets and one should choose e.g., a normalization method that make only minimal assumption about the presence of a set of constant miRNAs like invariant-based normalization (22). Alternatively, a normalization method free of assumption e.g., the majority of algorithms for variance stabilization normalization (41) or even an assumption free approach (42) can be utilized.

2.2.2.1 Quantile Normalization

Quantile normalization is a transformation method originally proposed by Bolstad et al. (39) for oligonucleotide arrays. It is now widely used for one-color miRNA microarrays as well and was confirmed as one of the most robust methods (21)(22)(35),(43). It is an inter-array approach and equalizes the distributions of expression intensities across arrays. Thus, quantile normalization assumes that the overall distribution of signal intensity does not change. While this assumption likely holds true for the comparison of p53 overexpressing versus control cells (22) or even for brainheart comparisons according to Rao et al. (35) where only 5% of miRNAs were differentially expressed, it may not hold true in case large numbers of miRNAs are differentially expressed in only one direction. Such cases may be e.g., knockouts of essential miRNA biogenesis proteins which lead to a dramatic reduction in steady state miRNA levels by blocking production of mature miRNAs (35). Rao at Al. in 2008 (35) compare the performance of several normalizations on miRNA single channel microarray profiling showing a better performance of quantile normalization.

2.2.2.2 LOESS Normalization

Between the transformation based methods, LOESS normalization and its variants (35)(21)(23) are the most used. They use local regression via locally weighted scatter plot smooth. It is advisable to introduce weights that penalize outliers because these values can strongly influence the local regression curve. Local regression via LOESS

uses a quadratic polynomial weighted regression function with Tukeys biweight function [(40) of the log ratios $\frac{Cy3}{Cy5}$ on overall spot intensity $Cy3 \times Cy5$ (the LOESS smoother for the so called MA-plots) (23). Hua et al. in 2008 (21) compared 15 normalization methods using microarray data and RT-PCR data. It was found that microRNA noramalized data by print-tip LOESS method were most consistent with the RT-PCR results. In addition, the two channel data normalization (using both Cy3 and Cy5 channels) is better than one-channel (using Cy3). Print-tip LOESS normalizes each M value by subtracting the corresponding value on the tip-group LOESS curve from the raw data (21). However, in a similar study, (24) did not find significant differences between print-tip LOESS and other normalizations. A variant of LOESS normalization called LOESSM was proposed by Risso et al. (23). This non-parametric normalization scales the expression data on the global median expression rather than on zero. This modification relaxes the assumption of symmetry among up- and down-regulated genes and it was shown that LOESSM, in case of absence of channel-effect, has better performance (23). In addition, LOESS combined with Generalized Procrustes Analysis (GPA)- an assumption free inter-array normalization (((42)) - improved its results and outperformed the other normalizations in terms of sensitivity and specificity (23). LOESS normalizations and its variants emerged as being robust in the reduction of non-biological bias.

2.2.2.3 Variance stabilization normalization

Variance stabilization normalization (VSN), an inter-array transformation method, is widely used for microRNA microarray data (24) (22). It was developed for mRNA arrays and is based on a parameterized arsinh transformation instead of a logarithmic transformation that calibrates sample-to-sample variations and renders variance approximately independent of the mean intensity (41). Spike-in VSN normalization as described restricts the model fit to spike-in spots. Normalization intensities for all miRNAs are then obtained by applying the resulting transformation to all spots of interest on the array (24). One limitations of this approach is that reliable results can only be obtained for intensities within the range covered by the spike-in used and that excludes targets that are not expressed. Pradervand et al. (22) proposed a linear regression method to select a set of miRNAs with constat expression (invariants) and used these invariants to calculate VSN parameter (VSN-INV). The invariant probes are those that have mediaum-high mean intesity and low variance across samples. VSN used with default parameter settings assumes that most genes are not differentially expressed whereas the invariant-based regression only assumes that a subpopulation of expressed genes does not change. So, VSN-INV is if a significant fraction of miRNAs is expected to be differentially expressed since, (22). Based on theirs comparisons, Pradervand et al. (22), found that VSN-INV and quantile normalization were the most robust normalization methods compared to VSN with default parameter or scaling. In general, one should note that VSN strongly affects the distribution of the large fraction of miRNAs whose expression is near or at background, resulting in the large increase of variability for those microRNAs (22).

2.2.2.4 Scaling normalization

The first normalization methods for mRNA microarray were based on the selections of predefined and stably expressed housekeeping genes as described by Garzon et al. (44) and Perkins at al. (45). Most commercially available miRNA microarrays do not have controls for endogenous RNAs that have been shown to be robustly invariant between various different tissue samples or conditions (35). To date, there is no consensus on the existence and reliability of reference gene miRNAs. The selection of reference genes to normalize miRNA levels depends on bioinformatic analysis of the respective data (as shown for mRNA (26)(30) and is otherwise still rather empirical due to the lack of robust reference miRNAs (46), although a universal reference miRNA reagent set has been proposed (24). Bargaje et al. (43) identified constitutively expressed miRNAs across tissues. The mean of expression levels of a set of 16 microRNAs showing minimum variability was reasonably successful as a normalization factor for comparing datasets generated by the same platforms. However, normalization using constitutive microRNAs was ineffective when comparing bead-based and microarray-based datasets. In these cases quantile and Z-score normalization were both successful in transforming the data sets generating comparable means and scale (43). The scaling methods like Z-score (43), mean, median (reviewed in (21)), or 75th percentile (20) assume that different sets of intensities differ by a constant global factor and all raw intensity values are multiplied with one common (i.e., global) scaling factor. The Z-score provides a mean-centered rank for the expression level in units of standard deviation. Z-scores thus provide an index of the expression level of the miRNA with respect to the cellular pool of miRNA. Unlike other normalization methods Z-scores are not influenced by the addition of new datasets allowing flexible cross-platform validation of miRNA microarray profiling experiments (43). Recently, Wang et al. (47) suggested the pre-evaluation of the overall miRNA expression pattern by a panel of miRNAs using RT-qPCR assays to build a logistic regression model based on these results. The personalized logistic regression model based on 29 miRNAs efficiently calibrated the variance across arrays and improved miRNA microarray discovery accuracy compared with different scaling methods, LOESS or quantile normalization (47).

2.3 Identification of differentially expressed genes and miRNA

Several methods have been applied to the identification of differentially expressed genes and microRNA in microarray data. The simplest method is to evaluate the log ratio between two conditions (or the average of ratios when there are replicates) and consider all the genes that differ by more than an arbitrary cut-off value to be differentially expressed. This is not a statistical test, and there is no associated value that can indicate the level of confidence in the designation of genes as differentially or not differentially expressed. It is considered to be unreliable (48) because statistical variability is not taken into account and is susceptible to outliers. More sophisticated statistical methods have been proposed. The classification success is affected by the choice of the method, the number of genes in the gene list, the number of cases (samples)

and the noise in the data set. Different methods produce dissimilar gene lists, which can produce dramatically different discrimination performance when trained as gene classifiers. The gene lists produced by the feature selection methods can be grouped broadly according to the manner in which they treat gene variance.

2.3.1 t-statistic

The simplest statistical method for detecting differential expression is t test. It can be used to compare two conditions when there is replication of samples. With more than two conditions, analysis of variance (ANOVA) can be used. The t-test calculates the observed t-statistic for each gene. The idea is to compare between-group difference and within-group difference and then to calculate the probability value (p-value) of t-statistic for each gene from t-distribution. The output of the analysis is a p-value for each gene. It represents the chance of getting the t-statistic as large as, or larger than the observed one, under the hypothesis of no differential expression (null hypothesis). A small p-value indicates that the hypothesis of no differential expression is not true and the gene is differentially expressed. The t-statistic methods perform relatively poorly when there are high levels of noise in microarray data together with low samples sizes. In this case, the variance estimate can be skewed by the genes which have a low variance. Due to the large numbers of genes present in microarray data sets, there will always be some genes which have a low standard deviation by chance. Thus, these genes will have a large t-statistic and will be falsely predicted to be differentially expressed.

2.3.2 SAM

Several modified t-statistics have been proposed to address this problem. SAM (49)61 is one of the most popular. It performs moderately well except when applied to data with low sample size and to the noisy data sets. SAM uses a moderated t-statistic, whereby a constant is added to the denominator of the t-statistic. The addition of this constant reduces the chance of detecting genes which have a low standard deviation by chance. The constant is estimated from the sum of the global standard error of the genes (50),(51),(52).

2.3.3 Empirical bayes method (Limma)

The empirical bayes method provides a more complex model of the gene variance. The gene standard error is estimated as a representative value of the variance of the genes at the same level of expression as the gene of interest (**?**) 65. In training sets with a large number of cases, the empirical bayes method performed comparably with ANOVA. Importantly, unlike most other methods, the empirical bayes t-statistic proved equally robust with low numbers of cases. The Bayesian statistic also provides p-values and has the advantage that it can be expanded to deal with datasets that have more then two classes. Limma provides advanced statistical methods for linear modelling of microarray data and for identifying differentially expressed genes. It fits a linear

model to the data and uses an empirical Bayes method for assessing differential expression (53). One or two experiment definition matrices need to be specified during the analysis: a design matrix defining the RNA samples and a contrast matrix (optional for simple experiments) defining the comparisons to be performed. When there are more than two conditions in an experiment, a more general concept of relative expression is needed. One approach that can be applied to cDNA microarray data from any experimental design is to use an analysis of variance model (ANOVA) to obtain estimates of the relative expression (VG) for each gene in each sample (54)(55). In the ANOVA model, the expression level of a gene in a given sample is computed relative to the weighted average expression of that gene over all samples in the experiment. The microarray ANOVA model is not based on ratios but it is applied directly to intensity data; the difference between two relative expression values can be interpreted as the mean log ratio for comparing two samples (as $logA - logB = log(\frac{A}{B})$), where logA and logB are two relative expression values). Alternatively, if each sample is compared with a common reference sample, one can use normalized ratios directly. This is an intuitive but less efficient approach to obtain relative expression values than using the ANOVA estimates. Direct estimates of relative expression can also be obtained from single-color expression assays (56). The set of estimated relative expression values, one for each gene in each RNA sample, is a derived data set that can be subjected to a second level of analysis. There should be one relative expression value for each gene in each independent sample. The distinction between technical replication and biological replication should be kept in mind when interpreting results from the analysis of a derived data. If inference is being made on the basis of biological replicates and there is also technical replication in the experiment, the technical replicates should be averaged to yield a single value for each independent biological unit. The derived data can be analyzed on a gene-by-gene basis using standard ANOVA methods to test for differences among conditions.

2.3.4 ROC

Classifiers built using gene lists from the ROC method outperform all other methods when applied to large datasets. High RCI scores are observed even when only a few of the most highly ranked genes are examined. These high RCI scores are maintained when the number of genes examined is increased. It is possible to obtain p-values using this method (57). ROC, like the t-statistic methods, loses power when the number of samples is reduced. It ranks a gene based on its power to discriminate between the groups given a threshold false positive rate. This means that it ignores the level of expression of the gene in the two groups. Therefore as the training size decreases, the likelihood of a gene with low variance and no biological meaning being a good discriminator by chance increases. ROC is an unsuitable method when the sample size is below 30 (class size of 15).

2.3.5 Rank Product

The Rank Product is a non-parametric method described in (52) and It generates a list of up- or down-regulated genes based on the estimated percentage of false positive predictions (pfp), which is also known as false discovery rate (FDR). The attractiveness of this method is its ability to analyse data sets from different origins (e.g. laboratories) or variable environments. Rank product assumes constant variance across all samples. It compares the product of the ranks of genes in a class with the product of the ranks of genes in the second class. For each gene in the data-set, rank products sorts the genes according to the likelihood of observing their ranked positions on the lists of differentially expressed genes just by chance.

2.4 Clustering

Clustering algorithms are widely used in the analysis of microRNA profiling data. In clinical studies, they are not only used to cluster microRNA into groups of co-regulated miRNA, but also for clustering patients, and thereby defying novel disease entities based on miRNA expression profiles. A reliable and precise classification of tumors is essential for successful diagnosis and treatment of cancer. Current methods for classifying human malignancies rely on a variety of morphological, clinical, and molecular variables. In spite of recent progress, there are still uncertainties in diagnosis. Also, it is likely that the existing classes are heterogeneous and comprise diseases which are molecularly distinct and follow different clinical courses. microRNA microarray datasets have been used to characterize the molecular variations among tumors by monitoring microRNA expression profiles on a genomic scale. This led to more reliable classification of tumors and to the identification of marker miRNA that distinguish among these classes. Eventual clinical implications include an improved ability to understand and predict cancer survival. However, there are three main types of statistical problems associated with tumor classification:

- The identification of new tumor classes using microRNA expression profiles unsupervised learning
- The classification of malignancies into known classes supervised learning
- The identification of marker microRNA that characterize the different tumor classes feature selection

Clustering can answer these problems. It is possible to cluster rows, columns or both. Rows (miRNA) clustering can identify groups of co-regulated miRNA, spatial or temporal expression patterns, reduce redundancy (cf. feature selection) in prediction, and detect experimental artifacts. On the other hand columns clustering allows to identify new classes of biological samples, new tumor classes or new cell types. Moreover, it allows to detect experimental artifacts. In order to perform clustering, a way to measure how similar or dissimilar two objects are is needed. The feature data are

often transformed to an $n \times n$ distance or similarity matrix, $D = d_{ij}$, for the n objects to be clustered. Features correspond to expression levels of different microRNAs and possible classes include tumor types or clinical outcomes (survival, non-survival). Other information such as age and sex may also be important and can be included in the analysis. The most popular distances are Euclidean distance and Manhattan distance. Hamming distance is used for ordinal, binary or categorical data. Clustering procedures can be divided into 3 categories: Hierarchical, Partitioning (K-means K-medoids/partitioning around medoids) and Model based approaches. The first one is either divisive or agglomerative and provides a hierarchy of clusters, from the smallest, where all objects are in one cluster, through to the largest set, where each observation is in its own cluster. One must often also dene a distance measure between clusters or groups of miRNA and the linkage methods used are single, complete, average, distance between centroids and Ward Linkage. Hierarchical clustering methods produce a tree or dendrogram. The partitions are obtained from cutting the tree at different levels. The tree can be built in two distinct ways bottom-up (agglomerative clustering) or top-down (divisive clustering). Examples of Hierarchical clustering methods are Self-Organizing Tree Algorithm SOTA (58) and DIvisive ANAlysis DIANA (59). Partitioning methods require the specification of the number of clusters. A mechanism for apportioning objects to clusters must be determined, and then data is portioned into a prespecied number K of mutually exclusive and exhaustive groups and iteratively reallocated to clusters until some criterion is met, e.g., minimize within-cluster sums-of-squares. Examples of partitioning methods are k -means and its extension to fuzzy k -means, Partitioning Around Medoids PAM, Self-Organizing Maps SOM and model-based clustering (59). An important feature of partitioning methods consists in satisfying an optimality criterion (approximately), however they need an initial K and long computation time. Hierarchical methods are computationally fast (for agglomerative clustering) but rigid, since they cannot later correct for earlier erroneous decisions. Most methods used in practice are agglomerative hierarchical methods. In large part, this is due to the availability of efficient exact algorithms that implement them. Model based approaches assume that data are 'generated' from a mixture of K distribution. They try to fit a model to the data and try to get the best fit. A classic example is a mixture of Gaussians (mixture of normals). They take advantage of probability theory and well-defined distributions in statistics. In microarray experiments is also useful to detect the presence of outliers. Outlier detection is an important step since they can greatly affect the between-cluster distances. Simple tests for outliers should be identifying observations that are responsible for a disproportionate amount of the within-cluster sum-of-squares. Most features in high dimensional datasets will be uninformative, examples are unexpressed genes, housekeeping genes, and 'passenger alterations'. Clustering (and classification) has a much higher chance of success if uninformative features are removed. Simple approaches to feature selection are: selecting intrinsically variable genes or requiring a minimum level of expression in a proportion of samples. Clustering can be also employed for quality control purposes. The clusters that are obtain from clustering samples/microRNA should be compared

with different experimental conditions such as batch or production order of the arrays, batch of reagents, microRNA amplification procedure, technician, plate origin of clones, and so on. Any relationships observed should be considered as a potentially serious source of bias.

2.5 Integrated analysis of miRNA and gene expression

miRNAs down-regulate their mRNA targets and this effect has shown to play a key role in different biological processes. miRNA regulatory mechanisms are complex and there is still no high-throughput experimental technique for miRNA target prediction. Although, in the last years several computational methods based on sequence complementarity of the miRNA and the mRNAs have been developed, their predictions are inconsistent and their expected false positive rates are large. Recently, new computational methods based on the joint analysis of miRNA and mRNA expression for the filtering of sequence-based putative targets have been proposed. Nevertheless, their expected false positive rates are still large and predictions of different methods do not match at all. Some of these methods combine both expression data with sequence analysis. The integration of miRNA and mRNA expression data have shown to be a good method for filtering sequence-based putative predictions. The algorithms to develop this integration can be categorized into three groups:

- dependence-based methods (Pearson and Spearman correlation and MI)
- MLR and regularized least squares (MLR, Lasso, Ridge and Elastic-net)
- Bayesian inference methods (GenMiR, HCtarget and a Bayesian graphicalmethod)

Although huge advances have been made in miRNA target prediction, there is still much work to do. Until high-throughput experimental techniques reach the market, computational methods will continue to be of high importance. Combination of expression data with sequence based prediction have shown to be feasible. Although, the number of predicted targets is still high, these methods have marked new future working lines. In this respect, models that combine more heterogeneous experimental data (i.e. TF, protein, time-course data, miRNA transfection effects on mRNA and proteins) could be more reliable on the predicted miRNAmRNA interactions.

2.6 Survival prediction model for cancer prognosis using gene expression

Cancer survival studies are commonly analyzed using survival-time prediction models for patients prognosis. Survival models consists of two parts: the underlying hazard function, describing how the hazard (risk) changes over time at baseline levels of covariates; and the effect parameters, describing how the hazard varies in response to explanatory covariates. The effect of covariates estimated by any proportional hazards

model can thus be reported as hazard ratios. The Cox proportional hazards model (60) is the most common survival prediction model for cancer prognosis. Sir David Cox observed that if the proportional hazards assumption holds (or, is assumed to hold) then it is possible to estimate the effect parameter(s) without any consideration of the hazard function. This approach to survival data is called application of the Cox proportional hazards model. Often, demographic and clinical covariates are combined in a Cox model to predict a patients survival in order to improve treatment recommendations (61) (62) (63). Many studies have shown an association between patient survival and gene expression profiles 8-10, thus some recent papers have investigated the use of microarray gene expression data alone or in combination with clinical covariates (64) (65) (66) as an improvement to estimate patient survival risk. Research in gene expression profiling of cancer data has focused on binary class prediction, where patients survival times have been dichotomized to form two classes (64) (67) (68). With this approach, a prediction model is built and used to distinguish between the low-risk and high-risk classes. Dimensionality reduction techniques are often performed prior to applying the Cox model to improve prediction performance. A practical approach is to select a smaller set of relevant genes from the entire gene set as initial step; a dimensionality reduction technique is then applied to the selected gene set (69). Evaluation of the ability of a survival model to predict future data is the most important consideration in the development of prediction model. The Hazard ratios between high- and low-risk groups defined by dichotomized risk scores are a common metric to assess the performance of survival prediction models. The KaplanMeier method (70) can be used to estimate survival curves for the two groups from the observed survival times without the assumption of an underlying probability distribution. The method is based on the basic idea that the probability of surviving k or more periods from entering the study is a product of the k observed survival rates for each period

$$S(k) = p_1 \times p_2 \times p_3 \times \ldots \times p_k$$

Here, p_1 is the proportion surviving the first period, p_2 is the proportion surviving beyond the second period conditional on having survived up to the second period, and so on. The proportion surviving period i having survived up to period i is given by:

$$p_i = \frac{r_i - d_i}{r_i}$$

Where r_i is the number alive at the beginning of the period and d_i the number of deaths within the period. Comparison of two or more survival curves can be done using a statistical hypothesis test known as log rank test (71). The null hypothesis of the test is that there is no difference between the population survival curves. The test statistic for two curves is calculated as follows:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

Where the O_1 and O_2 are the total numbers of observed events in groups 1 and 2,

respectively, and E_1 and E_2 the total numbers of expected events. The total expected number of events for a group is the sum of the expected number of events at the time of each event. The expected number of events at the time of an event can be calculated as the risk for death at that time multiplied by the number alive in the group. Under the null hypothesis, the risk of death (number of deaths/number alive) can be calculated from the combined data for both groups.

Chapter 3

Triple Negative Breast Cancer

3.1 Breast Cancer

After cardiovascular diseases, tumors represent the first cause of death in worldwide. Breast cancer is the most common type of malignancy diagnosed in the United States and Italy, after skin cancer. It is the second leading cause of cancer deaths in women today, after lung cancer. According to the American Cancer Society, more than 230,000 women will be diagnosed with breast cancer annually in the United States, and more than 39,000 will die from the disease. Official data from the Italian Ministry of Health have estimated the total breast cancer incidence at 37,300 new cases in year 2005, with an overall prevalence of 416,000 cases (women living with the cancer) [(72). The number of deaths due to breast cancer in the Italian female population represented about 18% of the total cancer mortality rate in 1998 and ten years later a total of 11.000 deaths were attributable to it. The risk of developing breast cancer is related to a number of factors including the events of reproductive life and lifestyle factors that modify endogenous levels of sex hormones (73). Diet has been also found to play an important role in the etiology of breast cancer (74). Tumors in the breast tend to grow slowly. By the time a lump is large enough to feel, it may have been growing for as long as 10 years. However, some tumors are aggressive and grow much more rapidly. Although breast cancer is often referred to as one disease, there are actually many different types of breast cancer. A major sub classification of breast cancer diving this tumor in invasive and non invasive (DCIS), is determined by the pathologist once he will analyze the tissue under microscope.

3.1.1 Invasive and Non-Invasive breast cancer

The Invasive breast cancer has spread from the original site (either the milk ducts or the lobules) into the surrounding breast tissue and possibly spread to the lymph nodes and/or other parts of the body forming metastasis. For this reason, invasive breast cancers have a poorer prognosis than DCIS. The invasive breast cancer can be further subdivided,and the most common is the invasive ductal carcinoma (also called

infiltrating ductal carcinoma and less commonly, invasive carcinoma of no special type or invasive carcinoma not otherwise specified). Invasive ductal carcinoma accounts for 50 to 75 percent of all breast cancers (75). Instead, Tubular carcinoma and mucinous (colloid) carcinoma are less common types of invasive breast cancer that tend to have a good prognosis (75). Ductal carcinoma *in situ* (DCIS) is a non-invasive breast cancer. In DCIS, the abnormal cells are contained in the milk ducts of the breast and have not spread into the surrounding breast tissue. Although DCIS is non-invasive, without treatment, the abnormal cells could turn into invasive breast cancer over time. That is way this is also called pre-invasive breast carcinoma to describe DCIS.

3.2 Molecular Subtype

If the previous sub classifications where mainly based on location and morphogenesis of the breast cancer cells and the tumor mass, comparing different types malignancy at the molecular level (mRNA, miRNA or protein) on a global scale could sub classify them resulting in one of the most striking discoveries of all the times; Subclassifing allowed the development of new therapies that save countless lives so far. Researchers have shown that microarray profiles could divide the breast cancers in different groups based, not on their morphology or location but on their dysregulated expressed genes. Perou et al. first identified distinct molecular sub types of breast cancer using unsupervised hierarchical clustering analysis of gene expression pattern differences (76). Similar classifications of breast cancers, using different unsupervised clustering analyses, have been seen by others (77),(78),(79). In 2009, a 50-gene signature (PAM50) was proposed to standardize breast cancer sub typing. The PAM50 gene set has high agreement in classification with larger 'intrinsic' gene sets previously used for sub typing (76)(80)(81). So far breast cancers are subdivided at the molecular scale because they differentially express fundamental proteins involved in tumor growth (but not only) like: ER (Estrogen Receptor), PR (progesterone Receptor), HER2 (Human Epidermal Growth Factor Receptor 2), Ki67 (Antigen KI-67)and EGFR 3.1:

Subtype	**These tumors tend to be**	**Prevalence**
Luminal A	ER+ and/or PR+, HER2-, low Ki67	40%
Luminal B	ER+ and/or PR+, HER2+ (or HER2- with high Ki67)	20%
Triple negative/basal-like	ER-, PR-, HER2-, cytokeratin 5/6 + and/or HER1+	15-20%
HER2 type	ER-, PR-, HER2+	10-15%

Table 3.1: Breast Cancer molecular sub-types

The ER is a member of the nuclear hormone family of intra cellular receptors, it is a DNA-binding transcription factor which regulates gene expression (82). There are two different forms of ER, referred as and , each encoded by a separate genes, the isoform by the ESR1 on chromosome 6 (6q25.1) and the isoform by the ESR2 gene on chromosome 14 (14q). These two forms of ERs are co-expressed in various tissues and they are over-expressed in around 70% of breast cancer cases, and are referred to as 'ER

positive'. Binding of estrogen to ER stimulates proliferation of mammary cells, with the resulting increase in cell division and DNA replication and increases mutation rate. This causes disruption of the cell cycle, apoptosis and DNA repair processes eventually leading to tumor formation.

The progesterone receptor (PR) is an intracellular steroid receptor that binds progesterone. PR is encoded by the PGR gene which lies on chromosome 11 (11q22). About 65% of ER-positive breast cancers are also PR-positive and about 5% of breast cancers are ER-negative and PR-positive. If cells have receptors for both hormones or receptors for one of the two hormones, the cancer is considered hormone-receptor positive. Co-regulators of PR either enhance or suppress transcription activity and thereby modulate its function.

The Human Epidermal growth factor Receptor 2 (HER2/neu or ERBB2) is a protein located at the long arm of chromosome 17 (17q11.2-q12). HER2/neu belongs to a family of four trans membrane receptor tyrosine kinases involved in signal transduction pathways that regulate cell growth and proliferation (83). Over-expression of this receptor in breast cancer is associated with increased disease recurrence and worse prognosis.

These three proteins are the main biomarkers used and besides them Ki-67 and EGFR (HER1) has become a very important predictive and prognostic marker for breast cancer. Antigen KI-67 (Ki-67) is a protein encoded by the MKI67 gene on chomosome 10. Ki-67 is a associated with and may be necessary for cellular proliferation; Patients with high Ki-67 expression responds better to chemotherapy (84)(85)(86)(87), but is associated with poor prognosis (88)(89)(90)(91). EGRF is a trans membrane glycoprotein that is a member of the protein kinase super family. It is a receptor for members of the epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor and has a key role in cell proliferation. Recent molecular profiling of these tumors has revealed a high frequency of its dysregulation, among other abnormalities. EGFR status correlates negatively with survival in patients with triple-negative breast cancers, and thus focus has turned on this receptor as a potential clinical target.

Researchers are now focusing their attention on finding new molecular breast cancer subtypes, based not only on different protein expression levels, but also on different miRNAs expression levels. This further classification is fundamental in order to develop more targeted treatment and therapies.

3.3 Triple negative breast

Triple-negative breast cancer (TNBC) patients are clinically negative for expression of estrogen and progesterone receptors (ER/PR) and HER2 protein. These are proteins that control cell functions, such as cell growth or death. Also, it seems to recur more often than other subtypes of breast cancer. It usually has a poorer prognosis than breast cancers that are hormone receptor-positive because lacks of specific, targeted treatment. TNBC is characterized by its unique molecular profile, aggressive behavior, distinct

patterns of metastasis, and lack of targeted therapies. Although not synonymous, the majority of triple-negative breast cancers carry the basal-like molecular profile on gene expression arrays. Basal-like tumors have cells with features similar to those of the outer (basal) cells lining the mammary ducts. Basal-like tumors tend to express HER1 and/or cytokeratin 5/6 proteins and most contain p53 mutations (92),(**?**). Most triple negative tumors are basal-like and most basal-like tumors are triple negative. However, not all triple negative tumors are basal-like and not all basal-like tumors are triple negative (as shown in the 3.1).

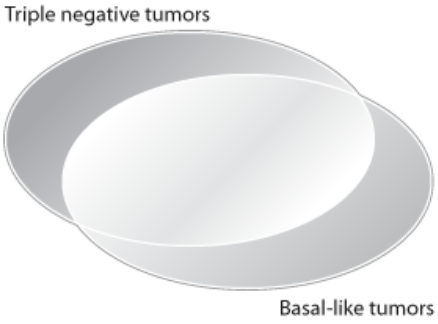

Figure 3.1: Venn Diagram between triple negative and basal-like tumors. - Triple-negative breast cancers are sometimes classified into "basal-type"; About 75% of basal-type breast cancers are triple negative.

About 15 to 20 percent of breast cancers are triple negative or basal-like (92),(93),(94),(95). Epidemiologic studies illustrate a high prevalence of triple-negative breast cancers among

- Younger women
- African descent
- Women who have BRCA1 mutations

Triple negative/basal-like tumors are often aggressive and have a poorer prognosis (at least within the first five years after diagnosis) compared to the estrogen receptor-positive subtypes (luminal A and luminal B tumors) (92)(93)(94)citeFan06. Although sensitive to chemotherapy, early relapse is common and a predilection for visceral metastasis, including brain metastasis, is seen. Targeted agents, including EGFR, vascular endothelial growth factor (VEGF), and poly (ADP-ribose) polymerase (PARP) inhibitors, are currently in clinical trials and hold promise in the treatment of this aggressive disease.

3.4 Integrated microRNA and mRNA Signatures Associated with Survival in Triple Negative Breast Cancer

Cascione L, Gasparini P, Lovat F, Carasi S, Pulvirenti A, Ferro A, Alder H, He G, Vecchione A, Croce CM, Shapiro CL, Huebner K. Paper submitted in PLoS One.

There is a major need to better understand the molecular basis of TNBC and to develop effective treatments for this aggressive type of breast cancer. More extensive genomic, molecular, and biological analyses of TNBCs are required to understand the complexity of the disease and to identify molecular drivers that can be therapeutically targeted. We compiled an extensive number of TNBC mRNA and microRNA profiles with the intent of linked specific miRNA signatures to patient survival and used miRNA anti-mRNA correlations to identify TNBC subclasses associated with expression of canonical signal pathways. We have used the nanoString nCounter platform (Seattle, WA, USA) to profile miRNA and mRNA expression in tumor, adjacent non-tumor (hereafter referred to as normal) and lymph node metastatic lesion (mets) tissues, from women with TNBCs; RNA was isolated from formalin-fixed paraffin-embedded (FFPE) tissue cores of 165 primary tumors, 59 adjacent normal and 54 lymph node metastatic samples and expression of 664 miRNAs and 230 cancer-associated mRNAs. The tissues studied are also represented on a tissue microarray with a database of associated clinical features and expression scores for some proteins that are differentially expressed in basal and non-basal TNBCs. Our analyses confirmed some observations from previous studies (96)(97) and revealed specific miRNA signatures as new potential biomarkers for distant-disease free (DDFS) and overall survival (OS). By multivariate analysis the risk signatures were independent predictors. We emphasized the joint analysis of miRNA and mRNA data, and analyzed correlations between miRNA and mRNA expression data. We show that particular cellular processes are significantly enriched in the co-regulated clusters, suggesting a central role for miRNAs in regulating these pivotal pathways. This study reveals specific miRNA expression profiles across the range of breast cancer patient-derived tissues. The identification of several molecular drivers provides great insight to the heterogeneity of this disease and provides preclinical platforms for the development of effective treatment.

Chapter 4

Research Papers

During my PhD program, I focused my research interests in the new and emerging research field of non coding RNAs and microRNAs. I focused my efforts on the investigating of the role of these short non coding RNAs in human tumors. During my PhD I had the opportunity of spending two years in Dr. Carlo M. Croce's lab at The Ohio State University in Columbus (USA).

As a member of Alfredo Ferro's Lab as well as member of Carlo Croce's Lab I have been involved in several projects. Many of them have been published in pear review journals. The following summary will illustrate how I maximized the excellent resources and opportunities available at the University of Catania and at The Ohio State University with a highly productive and sustainable research program.

4.1 Papers

Garofalo M, Romano G, Di Leva G, Nuovo G, Jeon YJ, Ngankeu A, Sun J, Lovat F, Alder H, Condorelli G, Engelman JA, Ono M, Rho JK, **Cascione L**, Volinia S, Nephew KP, Croce CM. ***EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers.*** Nature Medicine 2011.

In this project we aimed to identify EGFR- and MET-regulated miRNAs in non small cell lung cancer (NSCLC); I contributed to this work performing analysis, visualization and interpretation of high-throughput microRNA expression data. Appling bioinformatics and statistical methods we found that four miRNAs are modulated by both EGF and MET receptors, whereas two miRNAs are controlled only by MET. We showed that these miRNAs have important roles in gefitinib-induced apoptosis and epithelial-mesenchymal transition of NSCLC cells in vitro and in vivo. Our findings suggested that modulation of specific miRNAs may provide a therapeutic approach for the treatment of this particular kind of lung cancer. This paper was published in Nature Medicine

(98).

Ranganathan P, Heaphy CE, Costinean S, Stauffer N, Na C, Hamadani M, Santhanam R, Mao C, Taylor PA, Sandhu S, He G, Shana'ah A, Nuovo GJ, Lagana A, **Cascione L**, Obad S, Broom O, Kauppinen S, Byrd JC, Caligiuri M, Perrotti D, Hadley GA, Marcucci G, Devine SM, Blazar BR, Croce CM, Garzon R. ***Regulation of acute graft-versus-host disease by microRNA-155.*** Blood 2012.

In collaboration with members of Ramiro Garzons Lab. at The Ohio State University I had the opportunity to be be a part of this project. We investigated the miR-155 involvement in the modulation of acute graft-versus-host disease (aGVHD) the major complication of allogeneic hematopoietic stem cell transplant (alloHSCT). We found that miR-155 also affected aGVHD severity and prolonged survival in mice after alloHSCT. Our work discovered the role of miR-155 in the regulation of GVHD and pointed to miR-155 as a novel target for therapeutic intervention. In this work, I analyzed the data collected in the lab performed bioinformatics and statistical analyses. This paper was published in Blood (99).

Balatti V, Lerner S, Rizzotto L, Rassenti LZ, Bottoni A, Palamarchuk A, **Cascione L**, Alder H, Keating MJ, Kipps TJ, Pekarsky Y, Croce CM. ***Trisomy 12 CLLs progress through NOTCH1 mutations***. Leukemia 2012.

In this research, we investigated the molecular consequences of the constantly active expression of Notch1 protein in trisomy 12 B-cell chronic lymphocytic leukemia, one of the most common adult leukemia in Western societies. This gene has an important role in cell differentiation, proliferation and apoptosis, leading to transcriptional activation of multiple target genes, including MYC. Using Affymetrix microarray we compared the genome-wide mRNA expression of NOTCH1 Wild Type and NOTCH1-mutated samples. The clustering of samples showed a different expression pattern between the two categories (NOTCH1 WT and NOTCH1-mutated). The enriched analysis of dysregulated mRNAs reveled that the activation of NOTCH1 appears to be involved in downregulation of tumor suppressor and apoptotic key factors. accelerating the progression of the disease. The NOTCH1 mutations are associated with CLL progression leading to more aggressive form of the disease with poor outcome. This manuscript was published in Leukemia (100).

Lagana A, Paone A, Veneziano D, **Cascione L**, Gasparini P, Carasi S, Russo F, Nigita G, Macca V, Giugno R, Pulvirenti A, Shasha D, Ferro A, Croce CM.
miR-EdiTar: A database of predicted A-to-I edited miRNA target sites. Bioinformatics 2012.

The misexpression of microRNAs has been linked to altered cell behavior and the establishment and maintenance of malignant phenotypes (Croce 2009; Lagan et al. 2010; Sato et al. 2011). A-to-I editing is an important epigenetic mechanism that can

affect miRNA function by changing either their sequence or their binding sites on the targets (Nishikura 2010). Many of the edited sites are found in the mature miRNA sequences and some are located in the seed region, thus potentially altering target recognition. For example, a seed-edited version of miR-376 in mouse was proved to target a different set of genes than its unedited form (Kawahara 2007). We created miR-EdiTar, a database of predicted miRNA binding sites affected by A-to-I editing and novel binding sites genereted by this epigenetic mechanism. In this work we also provided a proof of principle validation of a novel miRNA binding site created by editing events and suggesting some plausible scenarios of the involvement of editing in miRNA activity. This paper was finally accepted for the publication in Bioinformatics (101).

Romano G, Acunzo M, Garofalo M, Di Leva G, **Cascione L**, Zanca C, Bolon B, Condorelli G, Croce CM. ***MiR-494 is regulated by ERK1/2 and modulates TRAIL-induced apoptosis in nonsmall-cell lung cancer through BIM down-regulation.*** PNAS 2012.

In this paper we accessed a functional relationship between the ERK1/2 pathway and BIM expression through miR-494. The ERK1/2 pathway has a key role in several cellular processes and cancer development and is responsible for the transcription of several important miRNAs. After ERK1/2 inactivation through the overexpression protein PED/PEA15, miR-494 was the most down-regulated microRNA . We alse found that this miRNA induced Tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) resistance in nonsmall-cell lung cancer (NSCLC) through the down-modulation of BIM. Elucidation of this undiscovered ERK1/2 pathway that regulates apoptosis and cell proliferation through miR-494 in NSCLC will greatly enhance our understanding of the mechanisms responsible for TRAIL resistance and will provide an additional arm for the development of anticancer therapies. This paper has been published in Proc. Natl. Acad. Sci. USA. (102).

Pichiorri F, Palmieri D, De Luca L, Consiglio J, You J, Rocci A, Talabere T, Piovan C, Lagana A, **Cascione L**,Guan J, Gasparini P, Balatti V, Coppola V, Hofmeister C, Marcucci G, Byrd J, Volinia S, Shapiro C, Freitas M, Croce CM. ***In vivo NCL-targeting affects breast cancer aggressiveness through miRNA regulation***. Journal of Exper. Medicine 2013.

In this project we found that the Nucleolin (NCL), a major nucleolar protein, post-transcriptionally regulates the expression of a specific subset of miRNAs (miR-21, miR-221, miR-222, and miR-103). These small non-coding RNA are causally involved in breast cancer initiation, progression and drug-resistance. We also shown that NCL is commonly overexpressed in human breast tumors, and its expression correlates with that of NCL-dependent miRNAs. Finally, this study indicates that NCL-binding guanosine-rich aptamers affect the levels of NCL-dependent miRNAs and their target genes, reducing breast cancer cell aggressiveness, both in vitro and in vivo. These findings illuminate a path to novel therapeutic approaches based on NCL-targeting

aptamers for the modulation of miRNA expression in the treatment of breast cancer. Our manuscript was accepted for the publication in Journal of Experimental Medicine (103).

4.2 Abstracts in Meeting

Rocci A, Hofmeister CC, Omed P, Geyer S, Bringhen S, **Cascione L**, Bingman A, Gambella M, Stiff A, Isaia G, De Luca L, Guan J, Rossi D, Corry J, Gentili S, Efebera Y, Uccello G, Benson DM, Ria R, Talabere T, Benevolo G, Murnan K, Callea V, Magarotto V, Boccadoro M, Croce CM, Palumbo A, Pichiorri F. ***Circulating microRNA in multiple myeloma: differences with healthy subjects and correlation with biological parameters.*** 17th Congress of the European Hematology Association, June 14-17 2012, Amsterdam - Netherlands.

The presence of microRNAs in many body fluid opens to their possible use as biomarkers. A comprehensive profile of circulating miRNA in multiple myeloma (MM) patients and a comparison with healthy subjects is still lacking. We identified the expression pattern of 654 miRNAs in 104 newly diagnosed MM patients and 60 from age-matched healthy subjects using nCounter technology (NanoString, Seattle, USA). We correlate the microRNAs expression profile with biological characteristics of multiple myeloma patients. Univariate analysis showed correlation between specific miRNAs and predictors of poor prognosis (ISS stage, FISH risk, high beta2-microglobulin values and low hemoglobin (Hb) levels). A diffuse reduction in miRNAs levels was observed in MM patients compared with healthy subjects and a low expression of specific miRNAs correlates with adverse prognostic factors in MM patients. Our observations strongly demonstrate a specific profile of circulating miRNAs in MM, opening the discussion on their role in the pathogenesis of the disease (104).

Shapiro CL, **Cascione L**, Gasparini P, Lovat F, Carasi S, Pulvirenti A, Ferro A, Huebner K. ***Use of microRNA (miR) expression profiling to identify distinct subclasses of triple-negative breast cancers (TNBC).*** 2012 American Society of Clinical Oncology Annual Meeting, June 1-5 2012, Chicago (IL) - USA.

To sub-classify TNBC we performed microRNA (miRNA) expression profiles using the nanoString nCounter platform) and linked them to patient overall survival. The consensus-clustering algorithm (ConsensusClusterPlus, Bioconductor) identified five distinct subclasses; 1 clustering with normal breast miRNA expression whereas the other 4 each had a unique pattern of deregulated miRNAs. The median overall survivals were significantly different across the 5 cancer subclasses (log-rank p=0.028). The miRNA expression profiling identifies and discriminates 5 TNBC subclasses, which do not coincide with those identified as basal and non-basal by IHC. Molecular analyses are ongoing to associate the miRNA-based subclasses with specific clinical features or

the expression of specific pathways (105).

Costinean S and Bottoni A, **Cascione L**, Teknos T, Ozer E, Old M, Agrawal A, Croce CM, Iwenofu OH. ***Differential microRNA Expression Signatures in Salivary Duct Carcinomas Versus Her2/Neu 3+ Positive Hormone Receptor Negative Invasive Ductal Breast Carcinomas and High Grade Breast Ductal Carcinoma In Situ.***United States and Canadian Academy of Pathology's 101st Annual Meeting, March 17-23 2012, Vancouver - Canada.

Salivary duct carcinoma (SDC) is highly lethal salivary gland tumor, histologically and immunohistochemically indistinguishable from invasive high grade Her2/Neu positive ductal breast carcinoma (IDBC). Herein, we sought to investigate whether common histopathologic and immunophenotypic features of the SDC and Her2/Neu 3+ positive IDBC have a similar molecular basis, in terms of microRNA expression. The expression patterns of Her2/Neu 3+ IDBC and SDC were strikingly similar and much less so with high grade DCIS. SDCs are more similar to Her2/neu 3+ IDBC than the high grade DCIS. Only two microRNAs were differentially expressed compared to IDBC and four microRNAs differentially expressed compared to the high grade DCIS. One of the microRNA differentially expressed - miR10a - was consistently higher in both high grade DCIS and Her2/Neu 3+ IDBC compared to the SDCs suggesting that this miRNA is breast specific and is increased from an early point of the tumorigenesis (106).

4.3 Chapter in book

Cascione L, Ferro A, Giugno R, Pigola G., Pulvirenti A. ***Algorithms and Methods for Expression Proling Data Classication and Biomarkers Identication.***Book Title: Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data, Wiley Book Series on Bioinformatics: Computational Techniques and Engineering

Microarray is a well established technology to analyze the expression of many genes and have recently become a basis for diagnosis and prognosis predictions in cancer research. In this chapter we first review the most reliable methods for expression profiling analysis and classification. We also present a comparative analysis on statistical tests on a case of study. Moreover, we present a new method for the computation of patterns of discriminant expressions genes for outcome prediction, called MIDClass. An experimental analysis show the effectiveness of this method compared to the most prominent classification approaches (107).

Cascione L, Ferro A, Giugno R, Pigola G., Pulvirenti A, Veneziano D. ***Elucidating the role of microRNAs in cancer through data mining techniques.*** Book Title: miRNA Cancer Regulation: Advanced Concepts, Bioinformatics and Systems Biology

Tools, Springer New York.

A new topic on bioinformatics miRNA research relies on the integration of heterogeneous data such as miRNA target predictions and expression profiles. It could be used to infer miRNA/phenotype associations and for the generation of network models of miRNA function. In this chapter we review the most important and recent methods for the analysis of miRNA expression profiles and the tools available on the web for functional analysis of miRNAs. Particular emphasis is given to the integration of heterogeneous data, including target predictions and expression profiles, which can be used to infer miRNA/phenotype associations and for the generation of network models of miRNA function. In particular, we describe the most used miRNA profiling technologies, together with the computational and statistical methods for the analysis of the related data. Emphasis is given to data normalization, the identification of differentially expressed microRNAs, clustering and the role of miRNAs as biomarkers. In addiction we given an overview of the most popular target prediction tools available on the web and finally we present a series of tools for functional analysis of miRNAs (108).

Chapter 5

Conclusion

5.1

To better understand the molecular basis of cancer, high throughput technologies are nowadays essentials. They generate a huge amount of biological data that needs computational and machine learning methods to handle it. The high throughput profiling expression studies have shown that miRNAs are dysregulated in a wide variety of human cancers (9) (10). In some instances, the expression of selected miRNAs or specific miRNA signatures was found to correlate with diverse clinico-pathological features and to predict patient clinical outcome and/or response to treatment (109) (110). Such findings have highlighted the potential of miRNAs as new diagnostic or prognostic/predictive biomarkers. Moreover, the role of miRNAs functioning as oncogenes and tumor suppressors, as emerged from functional studies in experimental models (111) (112) has generated great interest in their possible use as novel targets or tools for anticancer therapies. I make use of the microRNAs and genes profile to uncover biological relevant mechanisms that are disrupted or modified in the case of triple negative breast cancer. This study reveals also valuable insights in the prognosis of patients. We identified two microRNAs signatures able to predict the Distant Disease Free and Overall Survival. There is a substantial amount of original results that broadens our understanding of triple negative breast cancer and parts of the results confirm what has been previously published for TNBC. We also were able validated our results in three independent data sets. This kind of analysis is important to understand the molecular bases of human cancer, in order to develop treatments that can target specific molecular drivers. It is also crucial to identify specific bio-markers that will lead to early diagnosis of cancer. A better understanding of the role exerted by specific miRNAs in the development and progression of triple negative breast cancer is needed, as is a precise definition of their targets relevant to the disease. However, based on available findings, a possible role for miRNAs as novel bio markers and new therapeutic targets or intervention tools can be envisioned. This project leads to a substantial amount of results in the form of lists of differentially expressed microRNA, mRNA transcripts and affected protein-protein interaction network, which can be used as an inspiration for

a more targeted, experimental research. The microarray technology has now reached a mature state in terms of development, ease of use, costs and speed of analysis that makes it suitable for routine clinical use. Together with next generation sequencing, and other advanced data generating techniques, I believe that computer-aided diagnosis, personalized medicine and efficient data management must all be connected. The inter-disciplinarily of these research fields should be considered and promoted in order to have the best results that at the end will be reflected in terms of better life for cancer patients.

References

[1] D. Hanahan and R. A. Weinberg. **The hallmarks of cancer**. *Cell*, **100**(1):57–70, Jan 2000. 1, 2

[2] D. Hanahan and R. A. Weinberg. **Hallmarks of cancer: the next generation**. *Cell*, **144**(5):646–674, Mar 2011. 1, 2

[3] P. Anand, A. B. Kunnumakkara, A. B. Kunnumakara, C. Sundaram, K. B. Harikumar, S. T. Tharakan, O. S. Lai, B. Sung, and B. B. Aggarwal. **Cancer is a preventable disease that requires major lifestyle changes**. *Pharm. Res.*, **25**(9):2097–2116, Sep 2008. 3

[4] Erno Wienholds and Ronald H.A. Plasterk. **MicroRNA function in animal development**. *FEBS Letters*, **579**(26):5911 – 5922, 2005. ¡ce:title¿RNAi: Mechanisms, Biology and Applications¡/ce:title¿. 4

[5] C. M. Croce. **Causes and consequences of microRNA dysregulation in cancer**. *Nat. Rev. Genet.*, **10**(10):704–714, Oct 2009. 5

[6] G. Di Leva, D. Briskin, and C. M. Croce. **MicroRNA in cancer: new hopes for antineoplastic chemotherapy**. *Ups. J. Med. Sci.*, **117**(2):202–216, May 2012. 5, 10

[7] A. Cimmino, G. A. Calin, M. Fabbri, M. V. Iorio, M. Ferracin, M. Shimizu, S. E. Wojcik, R. I. Aqeilan, S. Zupo, M. Dono, L. Rassenti, H. Alder, S. Volinia, C. G. Liu, T. J. Kipps, M. Negrini, and C. M. Croce. **miR-15 and miR-16 induce apoptosis by targeting BCL2**. *Proc. Natl. Acad. Sci. U.S.A.*, **102**(39):13944–13949, Sep 2005. 15

[8] J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz, and T. R. Golub. **MicroRNA expression profiles classify human cancers**. *Nature*, **435**(7043):834–838, Jun 2005. 15

[9] S. Volinia, G. A. Calin, C. G. Liu, S. Ambs, A. Cimmino, F. Petrocca, R. Visone, M. Iorio, C. Roldo, M. Ferracin, R. L. Prueitt, N. Yanaihara, G. Lanza, A. Scarpa, A. Vecchione, M. Negrini, C. C. Harris, and C. M. Croce. **A microRNA expression signature of human solid tumors defines cancer gene targets**. *Proc. Natl. Acad. Sci. U.S.A.*, **103**(7):2257–2261, Feb 2006. 15, 43

[10] M. V. Iorio, M. Ferracin, C. G. Liu, A. Veronese, R. Spizzo, S. Sabbioni, E. Magri, M. Pedriali, M. Fabbri, M. Campiglio, S. Menard, J. P. Palazzo, A. Rosenberg, P. Musiani, S. Volinia, I. Nenci, G. A. Calin, P. Querzoli, M. Negrini, and C. M. Croce. **MicroRNA gene expression deregulation in human breast cancer**. *Cancer Res.*, **65**(16):7065–7070, Aug 2005. 15, 43

[11] G. A. Calin, C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini, and C. M. Croce. **Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers**. *Proc. Natl. Acad. Sci. U.S.A.*, **101**(9):2999–3004, Mar 2004. 15

[12] A. Ota, H. Tagawa, S. Karnan, S. Tsuzuki, A. Karpas, S. Kira, Y. Yoshida, and M. Seto. **Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma**. *Cancer Res.*, **64**(9):3087–3095, May 2004. 16

[13] A. Rouhi, D. L. Mager, R. K. Humphries, and F. Kuchenbauer. **MiRNAs, epigenetics, and cancer**. *Mamm. Genome*, **19**(7-8):517–525, Aug 2008. 16

[14] E. S. Lander. **The new genomics: global views of biology**. *Science*, **274**(5287):536–539, Oct 1996. 17

[15] F. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer. **A vision for the future of genomics research**. *Nature*, **422**(6934):835–847, Apr 2003. 17

[16] P. K. Kreeger and D. A. Lauffenburger. **Cancer systems biology: a network modeling perspective**. *Carcinogenesis*, **31**(1):2–8, Jan 2010. 18

[17] G. K. Geiss, R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, L. Hood, and K. Dimitrov. **Direct multiplexed measurement of gene expression with color-coded probe pairs**. *Nat. Biotechnol.*, **26**(3):317–325, Mar 2008. 18

[18] H. J. Peltier and G. J. Latham. **Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues**. *RNA*, **14**(5):844–852, May 2008. 19, 20

[19] K. H. Chang, P. Mestdagh, J. Vandesompele, M. J. Kerin, and N. Miller. **MicroRNA expression profiling to identify and validate reference genes for relative quantification in colorectal cancer**. *BMC Cancer*, **10**:173, 2010. 19, 20

[20] P. Lopez-Romero, M. A. Gonzalez, S. Callejas, A. Dopazo, and R. A. Irizarry. **Processing of Agilent microRNA array data**. *BMC Res Notes*, **3**:18, 2010. 19, 23

[21] Y. J. Hua, K. Tu, Z. Y. Tang, Y. X. Li, and H. S. Xiao. **Comparison of normalization methods with microRNA microarray**. *Genomics*, **92**(2):122–128, Aug 2008. 19, 20, 21, 22, 23

[22] S. Pradervand, J. Weber, J. Thomas, M. Bueno, P. Wirapati, K. Lefort, G. P. Dotto, and K. Harshman. **Impact of normalization on miRNA microarray expression profiling**. *RNA*, **15**(3):493–501, Mar 2009. 19, 20, 21, 22

[23] D. Risso, M. S. Massa, M. Chiogna, and C. Romualdi. **A modified LOESS normalization applied to microRNA arrays: a comparative evaluation**. *Bioinformatics*, **25**(20):2685–2691, Oct 2009. 19, 21, 22

[24] D. Sarkar, R. Parkin, S. Wyman, A. Bendoraite, C. Sather, J. Delrow, A. K. Godwin, C. Drescher, W. Huber, R. Gentleman, and M. Tewari. **Quality assessment and data analysis for microRNA expression arrays**. *Nucleic Acids Res.*, **37**(2):e17, Feb 2009. 19, 21, 22, 23

[25] F. Sato, S. Tsuchiya, K. Terasawa, and G. Tsujimoto. **Intra-platform repeatability and inter-platform comparability of microRNA microarray technology**. *PLoS ONE*, **4**(5):e5540, 2009. 19

REFERENCES

[26] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. *Genome Biol.*, **3**(7):RESEARCH0034, Jun 2002. 19, 20, 23

[27] A. Git, H. Dvinge, M. Salmon-Divon, M. Osborne, C. Kutter, J. Hadfield, P. Bertone, and C. Caldas. **Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression**. *RNA*, **16**(5):991–1006, May 2010. 19

[28] P. Mestdagh, P. Van Vlierberghe, A. De Weer, D. Muth, F. Westermann, F. Speleman, and J. Vandesompele. **A novel and universal method for microRNA RT-qPCR data normalization**. *Genome Biol.*, **10**(6):R64, 2009. 20

[29] M. W. Pfaffl, A. Tichopad, C. Prgomet, and T. P. Neuvians. **Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations**. *Biotechnol. Lett.*, **26**(6):509–515, Mar 2004. 20

[30] C. L. Andersen, J. L. Jensen, and T. F. ?rntoft. **Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets**. *Cancer Res.*, **64**(15):5245–5250, Aug 2004. 20, 23

[31] X. Wang. **A PCR-based platform for microRNA expression profiling studies**. *RNA*, **15**(4):716–723, Apr 2009. 20

[32] S. Baskerville and D. P. Bartel. **Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes**. *RNA*, **11**(3):241–247, Mar 2005. 20

[33] R. Q. Liang, W. Li, Y. Li, C. Y. Tan, J. X. Li, Y. X. Jin, and K. C. Ruan. **An oligonucleotide microarray for microRNA expression analysis based on labeling RNA with quantum dot and nanogold probe**. *Nucleic Acids Res.*, **33**(2):e17, 2005. 20

[34] H. Wang, R. A. Ach, and B. Curry. **Direct and sensitive miRNA profiling from low-input total RNA**. *RNA*, **13**(1):151–159, Jan 2007. 20

[35] Y. Rao, Y. Lee, D. Jarjoura, A. S. Ruppert, C. G. Liu, J. C. Hsu, and J. P. Hagan. **A comparison of normalization techniques for microRNA microarray data**. *Stat Appl Genet Mol Biol*, **7**(1):Article22, 2008. 20, 21, 23

[36] M. Chiogna, M. S. Massa, D. Risso, and C. Romualdi. **A comparison on effects of normalisations in the detection of differentially expressed genes**. *BMC Bioinformatics*, **10**:61, 2009. 20

[37] R. Garzon, F. Pichiorri, T. Palumbo, R. Iuliano, A. Cimmino, R. Aqeilan, S. Volinia, D. Bhatt, H. Alder, G. Marcucci, G. A. Calin, C. G. Liu, C. D. Bloomfield, M. Andreeff, and C. M. Croce. **MicroRNA fingerprints during human megakaryocytopoiesis**. *Proc. Natl. Acad. Sci. U.S.A.*, **103**(13):5078–5083, Mar 2006. 20

[38] E. A. Miska, E. Alvarez-Saavedra, M. Townsend, A. Yoshii, N. Sestan, P. Rakic, M. Constantine-Paton, and H. R. Horvitz. **Microarray analysis of microRNA expression in the developing mammalian brain**. *Genome Biol.*, **5**(9):R68, 2004. 20

[39] B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed. **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics*, **19**(2):185–193, Jan 2003. 20, 21

[40] C. Steinhoff and M. Vingron. **Normalization and quantification of differential expression in gene expression microarrays**. *Brief. Bioinformatics*, **7**(2):166–177, Jun 2006. 21, 22

[41] W. Huber, A. von Heydebreck, H. Sultmann, A. Poustka, and M. Vingron. **Variance stabilization applied to microarray data calibration and to the quantification of differential expression**. *Bioinformatics*, **18 Suppl 1**:96–104, 2002. 21, 22

[42] H. Xiong, D. Zhang, C. J. Martyniuk, V. L. Trudeau, and X. Xia. **Using generalized procrustes analysis (GPA) for normalization of cDNA microarray data**. *BMC Bioinformatics*, **9**:25, 2008. 21, 22

[43] R. Bargaje, M. Hariharan, V. Scaria, and B. Pillai. **Consensus miRNA expression profiles derived from interplatform normalization of microarray data**. *RNA*, **16**(1):16–25, Jan 2010. 21, 23

[44] R. Garzon, M. Garofalo, M. P. Martelli, R. Briesewitz, L. Wang, C. Fernandez-Cymering, S. Volinia, C. G. Liu, S. Schnittger, T. Haferlach, A. Liso, D. Diverio, M. Mancini, G. Meloni, R. Foa, M. F. Martelli, C. Mecucci, C. M. Croce, and B. Falini. **Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin**. *Proc. Natl. Acad. Sci. U.S.A.*, **105**(10):3945–3950, Mar 2008. 23

[45] D. O. Perkins, C. D. Jeffries, L. F. Jarskog, J. M. Thomson, K. Woods, M. A. Newman, J. S. Parker, J. Jin, and S. M. Hammond. **microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder**. *Genome Biol.*, **8**(2):R27, 2007. 23

[46] V. Benes and M. Castoldi. **Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available**. *Methods*, **50**(4):244–249, Apr 2010. 23

[47] B. Wang, X. F. Wang, P. Howell, X. Qian, K. Huang, A. I. Riker, J. Ju, and Y. Xi. **A personalized microRNA microarray normalization method using a logistic regression model**. *Bioinformatics*, **26**(2):228–234, Jan 2010. 23

[48] Y. Chen, E. R. Dougherty, and M. L. Bittner. **Ratio-based decisions and the quantitative analysis of cDNA microarray images**. *J Biomed Opt*, **2**(4):364–374, Oct 1997. 23

[49] V. G. Tusher, R. Tibshirani, and G. Chu. **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc. Natl. Acad. Sci. U.S.A.*, **98**(9):5116–5121, Apr 2001. 24

[50] S. Mukherjee, S. J. Roberts, and M. J. van der Laan. **Data-adaptive test statistics for microarray data**. *Bioinformatics*, **21 Suppl 2**:i108–114, Sep 2005. 24

[51] D. E. Martin, P. Demougin, M. N. Hall, and M. Bellis. **Rank Difference Analysis of Microarrays (RDAM), a novel approach to statistical analysis of microarray expression profiling data**. *BMC Bioinformatics*, **5**:148, Oct 2004. 24

[52] R. Breitling, P. Armengaud, A. Amtmann, and P. Herzyk. **Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments**. *FEBS Lett.*, **573**(1-3):83–92, Aug 2004. 24, 26

[53] M. K. Kerr, M. Martin, and G. A. Churchill. **Analysis of variance for gene expression microarray data**. *J. Comput. Biol.*, **7**(6):819–837, 2000. 25

[54] M. L. Lee, W. Lu, G. A. Whitmore, and D. Beier. **Models for microarray gene expression data**. *J Biopharm Stat*, **12**(1):1–19, Feb 2002. 25

[55] C. Li and W. Hung Wong. **Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application**. *Genome Biol.*, **2**(8):RESEARCH0032, 2001. 25

[56] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. **Exploration, normalization, and summaries of high density oligonucleotide array probe level data**. *Biostatistics*, **4**(2):249–264, Apr 2003. 25

[57] C. A. Tsai and J. J. Chen. **Significance analysis of ROC indices for comparing diagnostic markers: applications to gene microarray data**. *J Biopharm Stat*, **14**(4):985–1003, Nov 2004. 25

[58] J. Dopazo and J. M. Carazo. **Phylogenetic reconstruction using an unsupervised growing neural network that adopts the topology of a phylogenetic tree**. *J. Mol. Evol.*, **44**(2):226–233, Feb 1997. 27

[59] N. Semmar, B. Bruguerolle, S. Boullu-Ciocca, and N. Simon. **Cluster analysis: an alternative method for covariate selection in population pharmacokinetic modeling**. *J Pharmacokinet Pharmacodyn*, **32**(3-4):333–358, Aug 2005. 27

[60] E. F. Cox and D. Oakes. **Analysis of survival data**. *Chapman and Hall*, 1984. 29

[61] P. A. Gimotty, D. Guerry, M. E. Ming, R. Elenitsas, X. Xu, B. Czerniecki, F. Spitz, L. Schuchter, and D. Elder. **Thin primary cutaneous malignant melanoma: a prognostic tree for 10-year metastasis is more accurate than American Joint Committee on Cancer staging**. *J. Clin. Oncol.*, **22**(18):3668–3676, Sep 2004. 29

[62] X. Huang, S. Soong, W. H. McCarthy, M. M. Urist, and C. M. Balch. **Classification of localized melanoma by the exponential survival trees method**. *Cancer*, **79**(6):1122–1128, Mar 1997. 29

[63] M. Radespiel-Troger, W. Hohenberger, and B. Reingruber. **Improved prediction of recurrence after curative resection of colon carcinoma using tree-based risk stratification**. *Cancer*, **100**(5):958–967, Mar 2004. 29

[64] K. Shedden, J. M. Taylor, S. A. Enkemann, M. S. Tsao, T. J. Yeatman, W. L. Gerald, S. Eschrich, I. Jurisica, T. J. Giordano, D. E. Misek, A. C. Chang, C. Q. Zhu, D. Strumpf, S. Hanash, F. A. Shepherd, K. Ding, L. Seymour, K. Naoki, N. Pennell, B. Weir, R. Verhaak, C. Ladd-Acosta, T. Golub, M. Gruidl, A. Sharma, J. Szoke, M. Zakowski, V. Rusch, M. Kris, A. Viale, N. Motoi, W. Travis, B. Conley, V. E. Seshan, M. Meyerson, R. Kuick, K. K. Dobbin, T. Lively, J. W. Jacobson, and D. G. Beer. **Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study**. *Nat. Med.*, **14**(8):822–827, Aug 2008. 29

[65] A. Fernandez-Teijeiro, R. A. Betensky, L. M. Sturla, J. Y. Kim, P. Tamayo, and S. L. Pomeroy. **Combining gene expression profiles and clinical parameters for risk stratification in medulloblastomas**. *J. Clin. Oncol.*, **22**(6):994–998, Mar 2004. 29

[66] T. M. Habermann, S. S. Wang, M. J. Maurer, L. M. Morton, C. F. Lynch, S. M. Ansell, P. Hartge, R. K. Severson, N. Rothman, S. Davis, S. M. Geyer, W. Cozen, S. J. Chanock, and J. R. Cerhan. **Host immune gene polymorphisms in combination with clinical and demographic factors predict late survival in diffuse large B-cell lymphoma patients in the pre-rituximab era**. *Blood*, **112**(7):2694–2702, Oct 2008. 29

[67] J. Subramanian and R. Simon. **Gene expression-based prognostic signatures in lung cancer: ready for clinical use?** *J. Natl. Cancer Inst.*, **102**(7):464–474, Apr 2010. 29

[68] L. J. van 't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend. **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature*, **415**(6871):530–536, Jan 2002. 29

[69] E. Bair and R. Tibshirani. **Semi-supervised methods to predict patient survival from gene expression data**. *PLoS Biol.*, **2**(4):E108, Apr 2004. 29

[70] E. L. Kaplan and P. Meier. **Nonparametric estimation from incomplete observations**. *J. Amer. Statist. Assn*, **1**:457–481, 1954. 29

[71] N. Mantel. **Evaluation of survival data and two new rank order statistics arising in its consideration**. *Cancer Chemother Rep*, **50**(3):163–170, Mar 1966. 29

[72] E. Grande, R. Inghelmann, S. Francisci, A. Verdecchia, A. Micheli, P. Baili, R. Capocaccia, and R. De Angelis. **Regional estimates of breast cancer burden in Italy**. *Tumori*, **93**(4):374–379, 2007. 31

[73] T. Key, P. Appleby, I. Barnes, and G. Reeves. **Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies**. *J. Natl. Cancer Inst.*, **94**(8):606–616, Apr 2002. 31

[74] P. A. Baghurst and T. E. Rohan. **High-fiber diets and reduced risk of breast cancer**. *Int. J. Cancer*, **56**(2):173–176, Jan 1994. 31

[75] D.A. Dillon, A.J. Guidi, and S.J. Schnitt. **TO be defined**. *Int. J. Cancer*, **56**(2):173–176, Jan 1994. 32

[76] C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. L?nning, A. L. B?rresen-Dale, P. O. Brown, and D. Botstein. **Molecular portraits of human breast tumours**. *Nature*, **406**(6797):747–752, Aug 2000. 32

[77] C. Sotiriou, S. Y. Neo, L. M. McShane, E. L. Korn, P. M. Long, A. Jazaeri, P. Martiat, S. B. Fox, A. L. Harris, and E. T. Liu. **Breast cancer classification and prognosis based on gene expression profiles from a population-based study**. *Proc. Natl. Acad. Sci. U.S.A.*, **100**(18):10393–10398, Sep 2003. 32

[78] K. Yu, C. H. Lee, P. H. Tan, and P. Tan. **Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations**. *Clin. Cancer Res.*, **10**(16):5508–5517, Aug 2004. 32

[79] T. Sorlie, Y. Wang, C. Xiao, H. Johnsen, B. Naume, R. R. Samaha, and A. L. B?rresen-Dale. **Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms**. *BMC Genomics*, **7**:127, 2006. 32

[80] J. S. Parker, M. Mullins, M. C. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J. F. Quackenbush, I. J. Stijleman, J. Palazzo, J. S. Marron, A. B. Nobel, E. Mardis, T. O. Nielsen, M. J. Ellis, C. M. Perou, and P. S. Bernard. **Supervised risk predictor of breast cancer based on intrinsic subtypes**. *J. Clin. Oncol.*, **27**(8):1160–1167, Mar 2009. 32

REFERENCES

[81] Z. Hu, C. Fan, D. S. Oh, J. S. Marron, X. He, B. F. Qaqish, C. Livasy, L. A. Carey, E. Reynolds, L. Dressler, A. Nobel, J. Parker, M. G. Ewend, L. R. Sawyer, J. Wu, Y. Liu, R. Nanda, M. Tretiakova, A. Ruiz Orrico, D. Dreher, J. P. Palazzo, L. Perreard, E. Nelson, M. Mone, H. Hansen, M. Mullins, J. F. Quackenbush, M. J. Ellis, O. I. Olopade, P. S. Bernard, and C. M. Perou. **The molecular portraits of breast tumors are conserved across microarray platforms**. *BMC Genomics*, **7**:96, 2006. 32

[82] E. R. Levin. **Integration of the extranuclear and nuclear actions of estrogen**. *Mol. Endocrinol.*, **19**(8):1951–1959, Aug 2005. 32

[83] B. P. Zhou and M. C. Hung. **Dysregulation of cellular signaling by HER2/neu in breast cancer**. *Semin. Oncol.*, **30**(5 Suppl 16):38–48, Oct 2003. 33

[84] R. L. Jones, J. Salter, R. A'Hern, A. Nerurkar, M. Parton, J. S. Reis-Filho, I. E. Smith, and M. Dowsett. **Relationship between oestrogen receptor status and proliferation in predicting response and long-term outcome to neoadjuvant chemotherapy for breast cancer**. *Breast Cancer Res. Treat.*, **119**(2):315–323, Jan 2010. 33

[85] B. Keam, S. A. Im, H. J. Kim, D. Y. Oh, J. H. Kim, S. H. Lee, E. K. Chie, W. Han, D. W. Kim, W. K. Moon, T. Y. Kim, I. A. Park, D. Y. Noh, D. S. Heo, S. W. Ha, and Y. J. Bang. **Prognostic impact of clinicopathologic parameters in stage II/III breast cancer treated with neoadjuvant docetaxel and doxorubicin chemotherapy: paradoxical features of the triple negative breast cancer**. *BMC Cancer*, **7**:203, 2007. 33

[86] T. Petit, M. Wilt, M. Velten, R. Millon, J. F. Rodier, C. Borel, R. Mors, P. Haegele, M. Eber, and J. P. Ghnassia. **Comparative value of tumour grade, hormonal receptors, Ki-67, HER-2 and topoisomerase II alpha status as predictive markers in breast cancer patients treated with neoadjuvant anthracycline-based chemotherapy**. *Eur. J. Cancer*, **40**(2):205–211, Jan 2004. 33

[87] G. Pohl, M. Rudas, S. Taucher, T. Stranzl, G. G. Steger, R. Jakesz, R. Pirker, and M. Filipits. **Expression of cell cycle regulatory proteins in breast carcinomas before and after preoperative chemotherapy**. *Breast Cancer Res. Treat.*, **78**(1):97–103, Mar 2003. 33

[88] W. Weikel, C. Brumm, C. Wilkens, T. Beck, and P. G. Knapstein. **Growth fractions (Ki-67) in primary breast cancers, with particular reference to node-negative tumors**. *Cancer Detect. Prev.*, **19**(5):446–450, 1995. 33

[89] W. Weikel, T. Beck, M. Mitze, and P. G. Knapstein. **Immunohistochemical evaluation of growth fractions in human breast cancers using monoclonal antibody Ki-67**. *Breast Cancer Res. Treat.*, **18**(3):149–154, Aug 1991. 33

[90] S. Y. Jung, W. Han, J. W. Lee, E. Ko, E. Kim, J. H. Yu, H. G. Moon, I. A. Park, D. Y. Oh, S. A. Im, T. Y. Kim, K. T. Hwang, S. W. Kim, and D. Y. Noh. **Ki-67 expression gives additional prognostic information on St. Gallen 2007 and Adjuvant! Online risk categories in early breast cancer**. *Ann. Surg. Oncol.*, **16**(5):1112–1121, May 2009. 33

[91] R. W. Brown, C. D. Allred, G. M. Clark, C. K. Osborne, and S. G. Hilsenbeck. **Prognostic value of Ki-67 compared to S-phase fraction in axillary node-negative breast cancer**. *Clin. Cancer Res.*, **2**(3):585–592, Mar 1996. 33

[92] L. A. Carey, C. M. Perou, C. A. Livasy, L. G. Dressler, D. Cowan, K. Conway, G. Karaca, M. A. Troester, C. K. Tse, S. Edmiston, S. L. Deming, J. Geradts, M. C. Cheang, T. O. Nielsen, P. G. Moorman, H. S. Earp, and R. C. Millikan. **Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study**. *JAMA*, **295**(21):2492–2502, Jun 2006. 34

[93] D. C. Koboldt, R. S. Fulton, M. D. McLellan, H. Schmidt, J. Kalicki-Veizer, J. F. McMichael, L. L. Fulton, D. J. Dooling, L. Ding, E. R. Mardis, R. K. Wilson, A. Ally, M. Balasundaram, Y. S. Butterfield, R. Carlsen, C. Carter, A. Chu, E. Chuah, H. J. Chun, R. J. Coope, N. Dhalla, R. Guin, C. Hirst, M. Hirst, R. A. Holt, D. Lee, H. I. Li, M. Mayo, R. A. Moore, A. J. Mungall, E. Pleasance, A. Robertson, J. E. Schein, A. Shafiei, P. Sipahimalani, J. R. Slobodan, D. Stoll, A. Tam, N. Thiessen, R. J. Varhol, N. Wye, T. Zeng, Y. Zhao, I. Birol, S. J. Jones, M. A. Marra, A. D. Cherniack, G. Saksena, R. C. Onofrio, N. H. Pho, S. L. Carter, S. E. Schumacher, B. Tabak, B. Hernandez, J. Gentry, H. Nguyen, A. Crenshaw, K. Ardlie, R. Beroukhim, W. Winckler, G. Getz, S. B. Gabriel, M. Meyerson, L. Chin, P. J. Park, R. Kucherlapati, K. A. Hoadley, J. Auman, C. Fan, Y. J. Turman, Y. Shi, L. Li, M. D. Topal, X. He, H. H. Chao, A. Prat, G. O. Silva, M. D. Iglesia, W. Zhao, J. Usary, J. S. Berg, M. Adams, J. Booker, J. Wu, A. Gulabani, T. Bodenheimer, A. P. Hoyle, J. V. Simons, M. G. Soloway, L. E. Mose, S. R. Jefferys, S. Balu, J. S. Parker, D. Hayes, C. M. Perou, S. Malik, S. Mahurkar, H. Shen, D. J. Weisenberger, T. Triche, P. H. Lai, M. S. Bootwalla, D. T. Maglinte, B. P. Berman, D. J. Van Den Berg, S. B. Baylin, P. W. Laird, C. J. Creighton, L. A. Donehower, G. Getz, M. Noble, D. Voet, G. Saksena, N. Gehlenborg, D. DiCara, J. Zhang, H. Zhang, C. J. Wu, S. Y. Liu, M. S. Lawrence, L. Zou, A. Sivachenko, P. Lin, P. Stojanov, R. Jing, J. Cho, R. Sinha, R. W. Park, M. D. Nazaire, J. Robinson, H. Thorvaldsdottir, J. Mesirov, P. J. Park, L. Chin, S. Reynolds, R. B. Kreisberg, B. Bernard, R. Bressler, T. Erkkila, J. Lin, V. Thorsson, W. Zhang, I. Shmulevich, G. Ciriello, N. Weinhold, N. Schultz, J. Gao, E. Cerami, B. Gross, A. Jacobsen, R. Sinha, B. Aksoy, Y. Antipin, B. Reva, R. Shen, B. S. Taylor, M. Ladanyi, C. Sander, P. Anur, P. T. Spellman, Y. Lu, W. Liu, R. R. Verhaak, G. B. Mills, R. Akbani, N. Zhang, B. M. Broom, T. D. Casasent, C. Wakefield, A. K. Unruh, K. Baggerly, K. Coombes, J. N. Weinstein, D. Haussler, C. C. Benz, J. M. Stuart, S. C. Benz, J. Zhu, C. C. Szeto, G. K. Scott, C. Yau, E. O. Paull, D. Carlin, C. Wong, A. Sokolov, J. Thusberg, S. Mooney, S. Ng, T. C. Goldstein, K. Ellrott, M. Grifford, C. Wilks, S. Ma, B. Craft, C. Yan, Y. Hu, D. Meerzaman, J. M. Gastier-Foster, J. Bowen, N. C. Ramirez, A. D. Black, R. E. Pyatt, P. White, E. J. Zmuda, J. Frick, T. M. Lichtenberg, R. Brookens, M. M. George, M. A. Gerken, H. A. Harper, K. M. Leraas, L. J. Wise, T. R. Tabler, C. McAllister, T. Barr, M. Hart-Kothari, K. Tarvin, C. Saller, G. Sandusky, C. Mitchell, M. V. Iacocca, J. Brown, B. Rabeno, C. Czerwinski, N. Petrelli, O. Dolzhansky, M. Abramov, O. Voronina, O. Potapova, J. R. Marks, W. M. Suchorska, D. Murawa, W. Kycler, M. Ibbs, K. Korski, A. Spycha?a, P. Murawa, J. J. Brzezi?ski, H. Perz, R. ?a?niak, M. Teresiak, H. Tatka, E. Leporowska, M. Bogusz-Czerniewicz, J. Malicki, A. Mackiewicz, M. Wiznerowicz, X. V. Le, B. Kohl, V. T. Nguyen, R. Thorp, V. B. Nguyen, H. Sussman, D. P. Bui, R. Hajek, P. H. Nguyen, V. T. Tran, Q. T. Huynh, K. Z. Khan, R. Penny, D. Mallery, E. Curley, C. Shelton, P. Yena, J. N. Ingle, F. J. Couch, W. L. Lingle, T. A. King, A. M. Gonzalez-Angulo, G. B. Mills, M. D. Dyer, S. Liu, X. Meng, M. Patangan, F. Waldman, H. Stoppler, W. Rathmell, L. Thorne, M. Huang, L. Boice, A. Hill, C. Morrison, C. Gaudioso, W. Bshara, K. Daily, S. C. Egea, M. Pegram, C. Gomez-Fernandez, R. Dhir, R. Bhargava, A. Brufsky, C. D. Shriver, J. A. Hooke, J. L. Campbell, R. J. Mural, H. Hu, S. Somiari, C. Larson, B. Deyarmin, L. Kvecher, A. J. Kovatich, M. J. Ellis, T. A. King, H. Hu, F. J. Couch, R. J. Mural, T. Stricker, K. White, O. Olopade, J. N. Ingle, C. Luo, Y. Chen, J. R. Marks, F. Waldman, M. Wiznerowicz, R. Bose, L. W. Chang, A. H. Beck, A. M. Gonzalez-Angulo, T. Pihl, M. Jensen, R. Sfeir, A. Kahn, A. Chu, P. Kothiyal, Z. Wang, E. Snyder, J. Pontius, B. Ayala, M. Backus, J. Walton, J. Baboud, D. Berton,

M. Nicholls, D. Srinivasan, R. Raman, S. Girshik, P. Kigonya, S. Alonso, R. Sanbhadti, S. Barletta, D. Pot, M. Sheth, J. A. Demchok, K. R. Shaw, L. Yang, G. Eley, M. L. Ferguson, R. W. Tarnuzzer, J. Zhang, L. A. Dillon, K. Buetow, P. Fielding, B. A. Ozenberger, M. S. Guyer, H. J. Sofia, and J. D. Palchik. **Comprehensive molecular portraits of human breast tumours**. *Nature*, **490**(7418):61–70, Oct 2012. 34

[94] K. D. Voduc, M. C. Cheang, S. Tyldesley, K. Gelmon, T. O. Nielsen, and H. Kennecke. **Breast cancer subtypes and the risk of local and regional relapse**. *J. Clin. Oncol.*, **28**(10):1684–1691, Apr 2010. 34

[95] C. Fan, D. S. Oh, L. Wessels, B. Weigelt, D. S. Nuyten, A. B. Nobel, L. J. van't Veer, and C. M. Perou. **Concordance among gene-expression-based predictors for breast cancer**. *N. Engl. J. Med.*, **355**(6):560–569, Aug 2006. 34

[96] S. Valastyan, F. Reinhardt, N. Benaich, D. Calogrias, A. M. Szasz, Z. C. Wang, J. E. Brock, A. L. Richardson, and R. A. Weinberg. **A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis**. *Cell*, **137**(6):1032–1046, Jun 2009. 35

[97] B. D. Lehmann, J. A. Bauer, X. Chen, M. E. Sanders, A. B. Chakravarthy, Y. Shyr, and J. A. Pietenpol. **Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies**. *J. Clin. Invest.*, **121**(7):2750–2767, Jul 2011. 35

[98] M. Garofalo, G. Romano, G. Di Leva, G. Nuovo, Y. J. Jeon, A. Ngankeu, J. Sun, F. Lovat, H. Alder, G. Condorelli, J. A. Engelman, M. Ono, J. K. Rho, L. Cascione, S. Volinia, K. P. Nephew, and C. M. Croce. **EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers**. *Nat. Med.*, **18**(1):74–82, Jan 2012. 38

[99] P. Ranganathan, C. E. Heaphy, S. Costinean, N. Stauffer, C. Na, M. Hamadani, R. Santhanam, C. Mao, P. A. Taylor, S. Sandhu, G. He, A. Shana'ah, G. J. Nuovo, A. Lagana, L. Cascione, S. Obad, O. Broom, S. Kauppinen, J. C. Byrd, M. Caligiuri, D. Perrotti, G. A. Hadley, G. Marcucci, S. M. Devine, B. R. Blazar, C. M. Croce, and R. Garzon. **Regulation of acute graft-versus-host disease by microRNA-155**. *Blood*, **119**(20):4786–4797, May 2012. 38

[100] V. Balatti, S. Lerner, L. Rizzotto, L. Z. Rassenti, A. Bottoni, A. Palamarchuk, L. Cascione, H. Alder, M. J. Keating, T. J. Kipps, Y. Pekarsky, and C. M. Croce. **Trisomy 12 CLLs progress through NOTCH1 mutations**. *Leukemia*, Aug 2012. 38

[101] A. Lagana, A. Paone, D. Veneziano, L. Cascione, P. Gasparini, S. Carasi, F. Russo, G. Nigita, V. Macca, R. Giugno, A. Pulvirenti, D. Shasha, A. Ferro, and C. M. Croce. **miR-EdiTar: A database of predicted A-to-I edited miRNA target sites**. *Bioinformatics*, Oct 2012. 39

[102] G. Romano, M. Acunzo, M. Garofalo, G. Di Leva, L. Cascione, C. Zanca, B. Bolon, G. Condorelli, and C. M. Croce. **MiR-494 is regulated by ERK1/2 and modulates TRAIL-induced apoptosis in non-small-cell lung cancer through BIM down-regulation**. *Proc. Natl. Acad. Sci. U.S.A.*, **109**(41):16570–16575, Oct 2012. 39

[103] F. Pichiorri, D. Palmieri, L. De Luca, J. Consiglio, J You, A. Rocci, T. Talabere, C. Piovan, A. Lagana, L Cascione, J. Guan, P. Gasparini, V. Balatti, V. Coppola, C. Hofmeister, G. Marcucci, J. Byrd, S. Volinia, M. Shapiro, C. Freitas, and CM. Croce. **In vivo NCL-targeting affects breast cancer aggressiveness through miRNA regulation**. *J. Exp. Med.*, **209**(12):2137–2140, Nov 2012. 40

[104] A. Rocci, CC. Hofmeister, P. Omed, S. Geyer, S. Bringhen, L. Cascione, A. Bingman, M. Gambella, A. Stiff, G. Isaia, L. De Luca, J. Guan, D. Rossi, J. Corry, S. Gentili, Y. Efebera, DM. Uccello, G. Benson, R. Ria, T. Talabere, G. Benevolo, K. Murnan, V. Callea, V. Magarotto, M. Boccadoro, CM. Croce, A. Palumbo, and F. Pichiorri. **Circulating microRNA in multiple myeloma: differences with healthy subjects and correlation with biological parameters**. In *17th Congress of the European Hematology Association*. 40

[105] CL. Charles L. Shapiro, L. Cascione, P. Gasparini, F. Lovat, S. Carasi, A. Pulvirenti, A. Ferro, and K. Huebner. **Use of microRNA (miR) expression profiling to identify distinct subclasses of triple-negative breast cancers (TNBC)**. In *American Society of Clinical Oncology Annual Meeting*. 41

[106] S. Costinean, A. Bottoni, L. Cascione, T. Teknos, E. Ozer, M. Old, A. Agrawal, CM Croce, and Iwenofu OH. **Differential microRNA Expression Signatures in Salivary Duct Carcinomas Versus Her2/Neu 3+ Positive Hormone Receptor Negative Invasive Ductal Breast Carcinomas and High Grade Breast Ductal Carcinoma In Situ**. In *United States and Canadian Academy of Pathologys 101st Annual Meeting*. 41

[107] *Biological knowledge discovery handbook: preprocessing, mining and postprocessing of biological data*, chapter Microarray Data Analysis: From Preparation to Classification. Wiley, 2013. 41

[108] *MicroRNA Cancer Regulation*, chapter Elucidating the role of microRNAs in cancer through data mining techniques. Springer, 2013. 42

[109] R. Garzon, G. A. Calin, and C. M. Croce. **MicroRNAs in Cancer**. *Annu. Rev. Med.*, **60**:167–179, 2009. 43

[110] T. Sorlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning, and A. L. Borresen-Dale. **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc. Natl. Acad. Sci. U.S.A.*, **98**(19):10869–10874, Sep 2001. 43

[111] R. Garzon and C. M. Croce. **MicroRNAs and cancer: introduction**. *Semin. Oncol.*, **38**(6):721–723, Dec 2011. 43

[112] E. Sotillo and A. Thomas-Tikhonenko. **Shielding the messenger (RNA): microRNA-based anticancer therapies**. *Pharmacol. Ther.*, **131**(1):18–32, Jul 2011. 43