UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXVIII CICLO

Dottorato di Ricerca in
Scienze e Tecnologie Chimiche e Farmaceutiche

Studio della farmacogenomica e dei microRNA per il trattamento anticancro personalizzato

Pharmacogenetics of microRNAs for personalized anticancer treatment

Settore scientifico-disciplinare: **ING-IND/24**

Ph.D. STUDENT
Eva Dreussi

Ph.D. PROGRAM COORDINATOR
Professor Mauro Stener

THESIS SUPERVISORS
Professor Mario Grassi

Giuseppe Toffoli, MD

CO- SUPERVISOR
Erika Cecchin, PhD

ACADEMIC YEAR 2014 / 2015

Alla mia famiglia e a tutte le
persone importanti della mia vita

Questo lavoro di tesi è stato svolto presso la Scuola di Dottorato in Scienze e Tecnologie Chimiche e Farmaceutiche dell'Università degli Studi di Trieste, diretta dal Prof. Mauro Stener in collaborazione con la Struttura Operativa Complessa di Farmacologia Sperimentale e Clinica del Centro di Riferimento Oncologico di Aviano (Istituto di ricerca e cura a carattere scientifico), diretta dal Dott. Giuseppe Toffoli.

CONTENTS

ABSTRACT

New and innovative ways to conceive pharmacogenetic (PGx) studies are required to unravel the multitude of mechanisms that regulate patients' response to treatment and patients' prognosis. A broader approach in PGx can be realized with the introduction of new disciplines and tools, such as bioinformatics, and with a deeper investigation of the plethora of pathways that has not been included in PGx studies yet. Specifically, the analysis of those not strictly related with drug mechanism of action could open the way to new unexpected scenarios.

In this thesis we analyzed panels of polymorphisms (SNPs) potentially affecting microRNA (miRNA) maturation and immune system activity. We analyzed these pathways due to the important role they play in physiological and pathological conditions. In this thesis, we had the possibility to perform the genetic analyses with a medium-throughput technology, represented by BeadXpress. In addition, we enriched our study with the exploitation of bioinformatic tools at the beginning of the study, to select the genes and the SNPs to analyze, and at the end, to give some suggestions about the obtained genetic results.

We selected as clinical model the locally advanced rectal cancer (LARC). Specifically, the genetic analyses were performed to identify new biomarkers that can be translated in the personalization of neoadjuvant treatment and in the optimization of patients' follow-up.

One aim of this thesis was to define the capability of SNPs involved in miRNA activity and maturation to predict the complete pathological response to neoadjuvant treatment in terms of Tumour Regression Grade (TRG). To this purpose, we analyzed 114 SNPs on a group of 265 LARC patients homogeneously treated with fluoropyrimides-based chemoradiotherapy (CRT) in neoadjuvant setting. We identified five new potential predictive biomarkers of response to neoadjuvant treatment. *DROSHA*-rs10719 and *SMAD3*-rs17228212 were unfavorable predictive biomarkers (p=0.0274 and p=0.0049, respectively), while *SMAD3*-rs744910, *SMAD3*-rs745103, and *TRBP*-rs6088619 showed an opposite effect (p=0.0153, p=0.0471, and p=0.0125, respectively). The significant association of three SNPs of *SMAD3* and treatment response underlines its key role in this mechanism. Moreover, we performed a Classification And Regression Tree (CART) analysis to investigate the associations between genetic and clinico-pathological characteristics in affecting treatment response. We highlighted the importance of the time interval between the end of neoadjuvant therapy and surgery in defining treatment response. This result is really intriguing due to the possibility to modulate this clinical parameter.

Another central aim of this thesis was the identification of new potential prognostic biomarkers in a panel of 147 SNPs in immune-related genes analyzed in a group of 235 LARC patients. In particular, we tested their potential association with the 2-year disease free survival (2yDFS). The prognostic value of the significant SNPs were then tested with the overall survival (OS). At the end of this process, we identified three new prognostic biomarkers: *IL17F*-rs641701, *IL17F*-rs9463772, and *STAT3*-rs8069645 (p=0.003, p=0.002, and p=0.044, respectively). Additionally, we replicated the analyses on another group of 63 LARC patients who underwent radical surgery and adjuvant treatment based on fluoropyrimidines. The SNP *IL17F*-rs9463772 still resulted significant, highlighting its noteworthy prognostic role in this clinical model.

These analyses show some limits commonly related to PGx studies, as the possibility to select false positive biomarkers, the lack of a validation group of patients in the miRNA project, and the quite unknown functional role of the SNPs resulted significant in our study. Consequently, even if the obtained data are really fascinating, they claim for further analyses to fully validate the clinical value of the identified genetic variants.

To conclude, we strongly believe that a deeper study of pathways not strictly related with drug activity could be the key to a deeper comprehension of treatment response that ultimately affects patients' prognosis. We hope that all these efforts will have in the future a real clinical impact leading to an optimization of treatment planning and patients' follow-up.

1. RATIONALE

Many efforts have been made thus far in order to identify tools that can help clinicians to select the right drug for the right patient in the right moment, that is personalized medicine. Treatment tailoring represents a compelling need especially for drugs characterized by a low therapeutic index, such as those administered in cancer patients.

Pharmacogenetics (PGx) is aimed at the identification of biomarkers with potential predictive and prognostic value. This has also led to the introduction of some PGx analyses in the current clinical practice, as also reported by international guidelines. Nevertheless, considering the huge amount of literature available nowadays, only few PGx biomarkers have reached a real clinical impact. This can be ascribed to the fact that the majority of PGx studies has focused till now on well-known candidate pathways related to different treatments response, that represent, however, a multifactorial phenomenon that cannot be determined only by genes directly involved drug activity. Understanding treatment response can be of crucial importance considering that it is often related to patients prognosis. New and still unexplored scenarios need to be considered with a more comprehensive approach.

Regarding that, PGx studies have started to investigate new biological fields. One is represented by non coding RNAs (ncRNAs). They represent a field with big potentialities considering that, although over 90% of our genes encode for ncRNAs, the major part has not been studied yet. Moreover, it was extensively observed that microRNAs (miRNAs), a class of brief ncRNAs, can orchestrate cell destiny and have a strong impact also in pathological conditions, due to the key role they play governing hundreds of genes. The first literature evidences about the PGx of miRNAs have fuelled the interest about this field of study (Dreussi et al., 2012), although it needs to be further explored to clearly define the potential clinical value of these findings.

Another innovative topic is represented by immunogenetics, the study of the clinical impact of polymorphisms located in genes involved in immune activity. This can be explained considering the central role of immunity not only in tumorigenesis but also in treatment response (Grivennikov et al., 2010). In this case, the analysis of a system with a plethora of biological implications might be the key to fully understand a complex phenomenon such as treatment response.

PGx studies are evolving not only in the topics they are exploring but also in the methodologies they can exploit. A giant step has been made in this field thanks to the medium- and high throughput technologies available nowadays, that render possible the

acquisition of an unimaginable until recently amount of data. Another great advancement is represented by the introduction of bioinformatics, that can be harnessed both at the beginning of the study in order to select the genes and the polymorphisms (SNPs) of interest, and at the end to interpret, from a biological point of view, the obtained genetic results.

In this thesis, using a PGx approach, we have tried to face the clinical problem related to locally advanced rectal cancer (LARC) patients' management. According to the European guidelines, the treatment of this malignancy is based on neoadjuvant chemoradiotherapy (CRT) followed by surgery, and adjuvant chemotherapy, if needed (Glimelius et al, 2013). However, a general consensus about patients' management has not been reached. Specifically, two big problems arise:

- a complete response to neoadjuvant treatment is achieved only by a restricted portion of patients (Aschele et al., 2005; Probst et al., 2015). Thus, the identification of patients who have a low probability to respond to therapy can of great interest as they could avoid neoadjuvant treatment, sparing useless toxicities related to CRT and optimizing the time for the surgical intervention;
- the second problem is related to patients' prognosis. In particular, understanding who will relapse can be of help to clinicians. This can be ultimately translated into a more appropriate follow-up and, potentially, have an impact on the choice to treat patients in the adjuvant setting, that still represents an open question (Petrelli et al., 2015).

These two problems are of vital importance considering that the definition of patients' treatment has an enormous impact on their prognosis, that is the primary clinical dilemma. PGx analyses have already identified some potential genetic biomarkers (García-Flórez et al., 2015; Tan et al., 2011), that have not, however, reached the everyday clinical practice. Thus, a more comprehensive approach, less focused on the canonical pathways regulating PK and PD, could be of great interest to have new insights with a real clinical impact on rectal cancer treatment personalization. Moreover, the use of new medium- and high throughput technologies, as well as of bioinformatics tools, could facilitate and also ameliorate the potentialities of a PGx study about this malignancy.

2. AIMS

This thesis has the primary aim to explore new PGx fields in order to define predictive and prognostic biomarkers that can help clinicians to personalise rectal cancer patients' management. Specifically, during this PhD we have explored two innovative fields represented by SNPs potentially affecting miRNA activity and those involved in immune system activity. The potential clinical value of these SNPs were investigated in a homogeneous group of LARC patients.

In particular, this work aimed at:

1. defining procedures to select panels of SNPs involved in pathways of interest following new and innovative approaches based not only on literature analysis but also on the use of bioinformatics tools. In particular, we defined a procedure to select a panel of 144 SNPs potentially involved in miRNA maturation and activity and another panel of 192 SNPs affecting immune system. The analyses of two different pathways have required the use of different bioinformatics tools to achieve better and more specific results. These panels were optimized in order to be analyzable with a medium throughput methodology represented by BeadXpress;
2. identifying new potential predictive RT-dose independent genetic biomarkers of response to neoadjuvant treatment through the analysis of a group of LARC patients homogeneously treated in neoadjuvant setting. Specifically, the pathological tumour response (TRG) was selected as response parameter. Due to the complexity of this phenomenon, the secondary aim of this analysis was the definition of the interactions among the predictive genetic biomarkers and patients' clinical parameters known to be associated with response to neoadjuvant treatment;
3. identifying new potential prognostic biomarkers of LARC patients. In particular, we investigated the 2-year disease-free survival (2yDFS), the disease free survival (DFS) and the overall survival (OS). In the first part of this project we analyzed genetic biomarkers specifically associated with DFS. In the second part we moved our attention to the 2yDFS because it is a strong prognostic biomarker of OS and it may be used to identify LARC patients that need a more intensified adjuvant treatment (Valentini et al., 2015). To conclude, the primary aim was the identification of genetic biomarkers significantly associated with 2yDFS, and the secondary aim was to evaluate if the SNPs identified in the first step were also associated with 10-year OS. These step was applied to select biomarkers with a more robust prognostic value.

3. INTRODUCTION

"Physicians are having to become more acutely aware
of the unique circumstance of each patient-
something most people have long called for."
(Schork, 2015)

3.1 PERSONALIZED MEDICINE IN ONCOLOGY

"It's far more important
to know what person
the disease has
than what disease the person has."
(Hippocrates, ~460-370 BC)

When administering a treatment, clinicians hope to obtain effective responses without toxicities. Nonetheless, millions of people everyday take drugs that will not help them (*Figure 1)*: numerous studies have demonstrated that the obtained response rates can deeply vary among different therapeutic classes. In addition, 2.2 million adverse drug reactions (ADRs) occur each year in the US, including more than 100,000 deaths (Spear et al., 2001). In Europe, ADRs are responsible for a considerable amount of morbidity and mortality. According to a recently published study, 3.6% of all hospitalizations are due to ADRs, and up to 10% of patients in hospitals experience an ADR during their stay, which has resulted fatal in less than 0.5% of cases (Bouvy et al., 2015). To better monitor these events, international special programs for pharmacovigilance have started, guided both by the European Medical Agency (EMA) and by the American Food and Drug Administration (FDA).

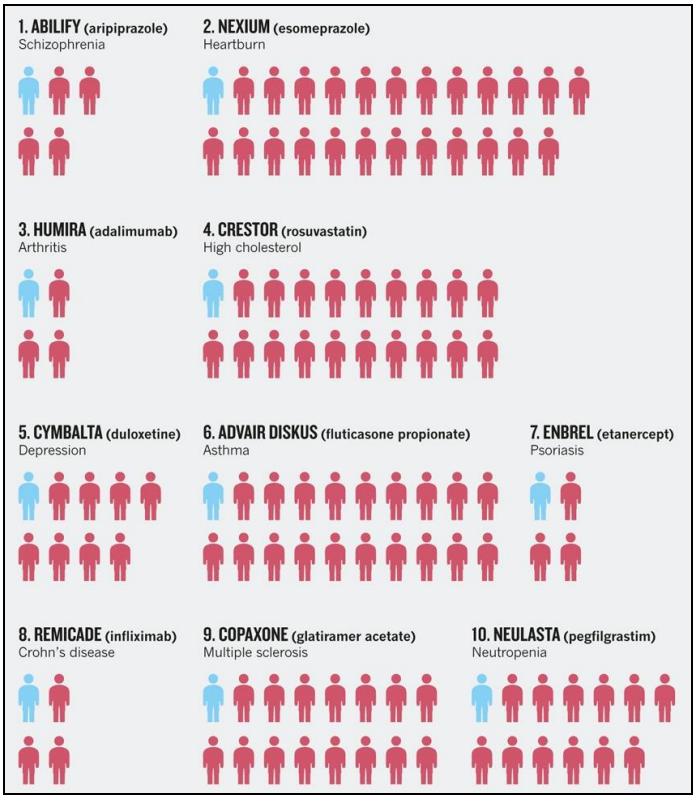

Figure 1: *Imprecision medicine*: the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people. In the figure, person who benefit from drugs are represented in blue, whereas persons who will not benefit are represented in red –adopted from (Schork, 2015)-.

The lack of the desired response becomes particularly serious considering drugs characterized by a low therapeutic index, such as those administered for cancer setting. A large part of administered treatment results indeed not only toxic but also ineffective: it is estimated that in only 25% of cases a response is achieved (Spear et al., 2001). Considering these data, many efforts have been made in order to personalize cancer patients' management, taking also advantage of the giant steps made in research both in the deepening of the biological mechanisms underlying tumours onset and in the technologies available nowadays. However, the majority of cancer patients is still waiting to reap the medical benefits of the post-genomic era. As well described elsewhere (Schork, 2015; Gravitz, 2014), new strategies that consider in a much integrated way the analysis of cancer genome, the potentialities of genetic engineering in oncology, the use of liquid biopsies to understand how tumour evolves, the application of nanodevices, and new ways to design clinical trials are requested.

Many parameters are responsible for the different responses observed in patients with the same diagnosis and treated with the same drugs, such as age, gender, comorbidities, dietary factors, lifestyle, and molecular background. Such observation has led clinicians and

researchers to change the way to conceive patients, highlighting the uniqueness of each clinical case that paved the way to the so called era of "*personalized medicine*". Actually, this concept dates back many hundreds of years with Hippocrates but only now, with the possibility to sequence the entire genome and the enormous evolution in computational biology and in other medical areas, we start to realize it. As reported by the FDA (Strimbu and Tavel, 2010; FDA 2013), many definitions of this term exist, and in particular we can mention these ones:

- ✓ "*The use of new methods of molecular analysis to better manage a patient's disease or predisposition to disease*" (Personalized Medicine Coalition)
- ✓ *"Providing the right treatment to the right patient, at the right dose at the right time"* (European Union)
- ✓ *"The tailoring of medical treatment to the individual characteristics of each patient"* (President's Council of Advisors on Science and Technology)
- ✓ *"Health care that is informed by each person's unique clinical, genetic, and environmental information"* (American Medical Association)
- ✓ *"A form of medicine that uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease"* (National Cancer Institute, NIH)

All these definitions give heavy importance to the individual peculiarity in terms of clinical-, genetic-, and environmental information, factors that can impact disease risk, treatment response, and patients' prognosis. This becomes essential in the oncology field, where a big concern due to toxicities and efficacy of the employed drugs exists.

Till now, the main goal obtained in the translation of the concept of personalized medicine in clinics can be ascertained to the identification of biomarkers, which can be defined as "*any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease*" (Strimbu and Tavel, 2010). They are broadly classified as diagnostic, predictive, and prognostic biomarkers, according to their clinical function. For instance, the identification of the Philadelphia chromosome responsible of the Bcr-Abl protein fusion in the chronic myeloid leukemia represents a watershed in leukemic patients' treatment, thanks to the introduction in the clinical practice of Imatinib, a drug specifically targeting the fusion protein.

Despite all these extraordinary advances, understanding why different individuals respond differently to treatment seems yet a distant goal. This means that clinicians have no choice but to follow a "trial and error" approach in prescribing treatments. To try to overcome this problem, the American President Barack Obama launched in January 2015 the national *Precision Medicine Initiative*, founding it with 215 million dollars, aiming at promoting the

introduction of personalized medicine concepts into the clinical practice, with a special focus on oncology and genetics.

However, in order to translate genomics into clinics it is necessary to face complex challenges. We can refer for instance to the difficulty related to the management of the magnitude of genetic information that we can obtain nowadays: which tools can record all these data and can be used to manage them properly? How to deal with the ethical issues related to genetics? How to ensure that legislation, clinicians, infrastructure can evolve to properly introduce into clinics tools involved in personalized medicine? Which is the real clinical validity of the genetic test? When is it necessary to revise labeling of already approved treatments according to the obtained new information? Moreover, after having verified the clinical validity of genetic information, it is also necessary to develop adequate tools to perform the laboratory test. This has led to some examples of successful co-development of target therapies and the relative mutation test. For instance, FDA has approved in 2011 simultaneously vemurafenib, that inhibits BRaf V600E, and the corresponding mutation test, Cobas® 4800 BRAF V600 Mutation Test. This mutation is found in 50% melanoma patients, so the genetic test is necessary to identify patients who will really benefit from the drug.

Finding an answer to all the aforementioned questions goes beyond the scope of this thesis. However, it is necessary to keep in mind all these scientific, technical, ethical, economic, and legal issues when speaking about the translation of genetic information into clinical practice.

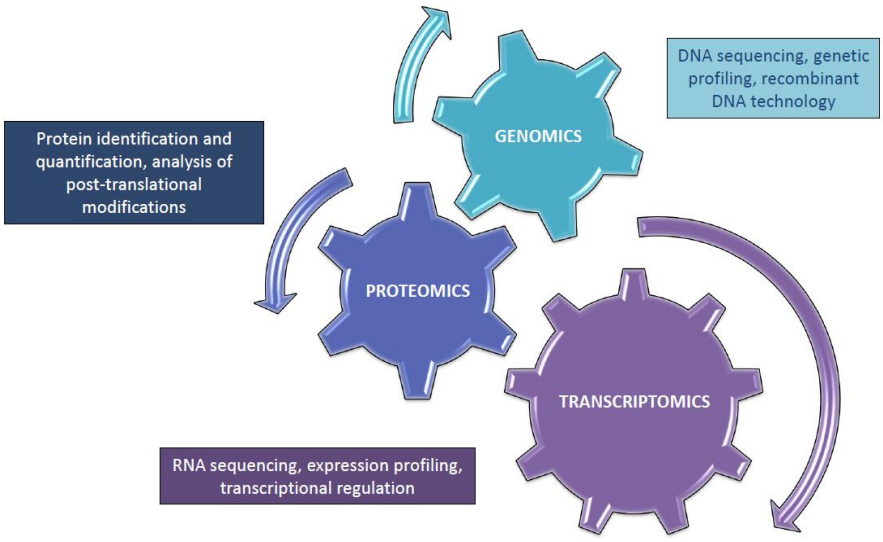

Figure 2: "*Omics era*": the main "-omics" disciplines related with molecular biology in personalized medicine.

To conclude, it is necessary to underline that not only genetics but also other scientific branches, like proteomics, metabolomics,… are identifying new biomarkers with high potential clinical value (*Figure 2*). We have entered the so called "*omics era*" and its impact on the clinical daily practice will be clarified in the next years.

3.2 PHARMACOGENOMICS

One pivotal science in the personalized medicine era is represented by pharmacogenomics (PGx). It aims at defining genetic predictive and prognostic biomarkers that can help clinicians in selecting the more appropriate treatment for each patient. It studies common variations in DNA sequences, called polymorphisms, like nucleotides' deletion/insertion, microsatellites, and single nucleotide polymorphisms (SNPs). The relevance of these variations is related to their possible implication in gene expression regulation or in their influence on protein activity, thus impacting pharmacokinetics (PK) and pharmacodynamics (PD). As a consequence, once identified variations with clinical value, patients can be stratified according to their genotype and more appropriate treatments can be selected.
Some PGx biomarkers have been already introduced in drug labeling, both in Europe and USA. Currently, more than 100 approved drugs contain information about genomic biomarkers (FDA 2013). In *Table 1*, a list of drugs administered to cancer patients, approved by FDA and with a pharmacogenomic recommendation is reported. The most updated information related to PGx can be found on *PharmGKB-The Pharmacogenomics Knowledgebase*.

DRUG	GENE
Ado-Trastuzumab emtansine	*ERBB2*
Afatinib	*EGFR*
Anastrozole	*ESR1, PGR*
Arsenic Trioxide	*PML-RARA*
Bosutinib	*BCR/ABL1*
Busulfan	*BCR-ABL1*
Capecitabine	*DPYD*
Ceritinib	*ALK*
Cetuximab	*EGFR, KRAS*
Cisplatin	*TPMT*
Crizotinib	*ALK*
Dabrafenib	*BRAF, G6PD*
Dasatinib	*BCR/ABL1*
Denileukin Diftitox	*IL2RA*
Erlotinib	*EGFR*
Everolimus	*ERBB2, ESR1*
Exemestane	*ESR1, PGR*
Fluorouracil	*DPYD*
Fulvestrant	*ESR1, PGR*
Ibrutinib	*del (17p)*
Imatinib	*KIT, BCR-ABL1, PDGFRB, FIP1L1-PDGFRA*
Irinotecan	*UGT1A1*
Lapatinib	*ERBB2, HLA-DQA1, HLA-DRB1*
Letrozole	*ESR1, PGR*
Mercaptopurine	*TPMT*
Nilotinib	*BCR-ABL, UGT1A1*
Obinutuzumab	*MS4A1*
Omacetaxine	*BCR-ABL1*
Panitumumab	*EGFR, KRAS*
Pazopanib	*UGT1A1*
Pertuzumab	*ERBB2*
Ponatinib	*BCR-ABL1*
Rasburicase	*G6PD, CYB5R1-4*
Rituximab	*MS4A1*
Tamoxifen	*ESR1, PGR, F5, F2*
Thioguanine	*TPMT*
Tositumomab	*MS4A1*
Trametinib	*BRAF*
Trastuzumab	*ERBB2*
Tretinoin	*PML/RARA*
Vemurafenib	*BRAF*

Table 1: *FDA-approved drugs in oncology with pharmacogenomic information in their labeling*: biomarkers in the table include germline and somatic gene variants, functional deficiencies, expression changes, chromosomal abnormalities, and protein biomarkers. Table downloaded from FDA website (last page update: 20/05/2015).

Many classifications of the genetic variants exist, according to their effect on mRNA or on proteins, the alterations they are responsible for in the DNA sequence, etc.

In this thesis we have analyzed SNPs, common single base changes naturally occurring in DNA sequence. The official catalog of nucleotide changes in human and other organisms is dbSNP, held by the National Center for Biotechnology Information (NCBI). To date, according to NCBI dbSNP Build 144, 161.454.519 genetic variants have been annotated. As already stated, these variants could affect genetic transcription and also protein activity. Nonetheless, it is astonishing that we are often not able to associate a polymorphism with its biological function, rendering the interpretation of genetic data more puzzling.

The main approaches that can be selected when planning a PGx study are the following ones:

- *Gene-candidate study*: genes are chosen based on their physiological or pharmacologic effect on disease or drug response, thus having a high probability to obtain good and understandable results. This approach is cost-effective and requires a small sample size. The main drawback of these studies is that it is necessary to know before the gene function, and, as previously said, defining the impact of a genetic variant on drug response is often not so linear.
- *Pathway-candidate study*: the study focuses on genes involved in a crucial pathway regulating the disease onset or the PK or PD. A prior knowledge of the selected pathway is necessary, even if the interpretation of the results can be more complicated. One single gene is indeed involved in different pathways, so it is quite complicated to specifically define the real impact of the selected pathway on the process of interest. On the other hand, this approach can broaden the initial hypothesis shedding light to new and unexpected scenarios.
- *Genome wide association studies (GWAS)*: technologies available nowadays make possible the study of panels of 10^5-10^6 genetic variations. With this approach you can highlight new and unexpected associations between genetic information and the clinical problem you are investigating. Nevertheless, the study of huge numbers of SNPs requires a robust statistical analysis and a very large sample size to overcome problems related to false positive biomarkers. Moreover, this process is time consuming, expensive, and demands suitable technological platforms and qualified staff to perform these analyses and to correctly interpret the obtained results.

PGx is evolving under different aspects due to the giant steps that have been made from both a technical and a scientific point of view. In the next paragraphs, a brief overview of PGx is described. In particular, in the first paragraph two new fields of study are presented: the PGx of microRNAs (miRNAs) and immunogenetics. Secondly, we will briefly outline the

technological advancements in PGx, highlighting in particular the tools exploited in this thesis.

3.2.1 Evolution of pharmacogenetics : new fields of study

Beyond protein -coding regions : non coding RNAs (ncRNAs)

The scarce clinical translation of the PGx markers can be addressed considering different limits of these analyses, like the lack of a proper study design, of rigorous statistical models, and of large and homogeneous groups of patients to analyze as training and validation set. Moreover, the majority of the PGx studies have focused till now on well-known candidate pathways related to different treatments response, but gene expression is a multifactorial phenomenon dependent not only on genetic factors, but also on epigenetic mechanisms, that need to be better elucidated and that maybe in the future will help clinicians to ameliorate patients' treatment. Specifically, in the last years non protein coding RNAs (ncRNAs) have reached the attention of researchers due to the pivotal role they play in gene expression control.

It is estimated that over 90% of our genes encode for ncRNAs. Various structural and functional classes of ncRNAs have been described, and in many cases their functional role in cellular physiology has also been elucidated. In particular, many efforts have been spent to study the role of small ncRNAs, even if in the last years much interest was also invested to understand the role of long ncRNAs. Among them, one of the most studied is HOTAIR: its prognostic role in different cancers like colorectal cancer (CRC)(Kogo et al., 2011) and glioma (Zhang et al., 2013) has been already defined.

Small ncRNAs are defined as products with a length of less than 400 nucleotides. They are involved in gene control acting at different levels: transcriptional silencing, transcriptional activation, alternative splicing. These processes intertwine with other ones regulating DNA methylation, histone modification, and chromatin remodeling, creating a complex network that finely tunes gene expression (Martens-Uzunova et al., 2013).

Class of ncRNA	Length (nt)	Functions	Cancer-related ncRNAs
small nucleolar RNA (snoRNA)	60-300	-post-transcriptional modifications of RNAs -nucleoside modifications (2'O-ribose methylation and pseudouridylation)	ACA11 - overexpressed in bladder-, colon-, and esophageal cancers - associated with resistance to doxorubicin, cellular response to stress, and proliferation of multiple myeloma cells
Piwi interacting RNA (piRNA)	26-31	-genome integrity, transposon repression, epigenetics - germline development	piR-651 -gastric cancer piR-823 -gastric cancer and multiple myeloma -regulation of angiogenesis and DNA methylation piR-Hep1 -hepatocarcinoma -involved in cellular proliferation, motility, and invasion
Vault RNA (vRNA)	~100	-components of the complex with Vault protein -possibly involved in multidrug resistance phenomena -possibly involved in miRNA generation	vRNA1-1 and vRNA1-2 -resistance to mitoxantrone vRNA2-1-5p -resistance to cisplatin
microRNA (miRNA)	19-25	-mRNA degradation and inhibition of mRNA translation -other mechanisms of translational and transcriptional control	LCS6 (SNP on *K-Ras* affecting Let7 binding) -response to neoadjuvant treatment in rectal cancer - response to docetaxel in non small cell lung cancer premiR-499 (rs3746444) -response to chemotherapy in advanced gastric cancer

Table 2: *Classes of small ncRNAs*: classes of small ncRNAs discussed in this thesis are reported in this table along with their length, the functions they physiologically exert inside the cell as well as the members of small ncrnas involved in cancer. regarding miRNA, we reported only some examples. a more comprehensive analysis is reported in the specific paragraph. ncRNA: protein non coding RNA; nt: nucleotide.

In *Table 2* the different classes of short ncRNAs described in this thesis are reported. In the next paragraph, the main small classes of ncRNAs are briefly described, with a special focus on their role in cancer biogenesis and, if known, in response treatment. To follow, there is a paragraph on a special class of small ncRNAs that we have investigated in this thesis, the microRNAs (miRNAs). The presentation is particularly oriented in the description of treatment response biomarkers.

Small nucleolar RNAs (snoRNAs)

Small nucleolar RNAs (snoRNAs) form a large family of well-conserved small ncRNAs, 60-300 nucleotides in length, involved in ribosome biogenesis. They are present in all

eukaryotes. In vertebrates, almost all snoRNAs are encoded in introns of host genes, even if it they can also act as mobile genetic elements and copy themselves as retrotransposons. They are usually transcribed with the host gene and generated during the splicing process. Mature snoRNAs localize in a specific nuclear compartment, the nucleolus, where they regulate the post-transcriptional modification of ribosomal RNAs and of other small nuclear RNAs. Moreover, they form complexes with specific ribonucleoproteins.

According to their sequence motifs and to their function, snoRNAs can be roughly divided in 3 classes:

- the *box C/D* snoRNAs, that guide 2'O-ribose methylation of target RNAs;
- the *box H/ACA*, that promote the pseudouridylation;
- the *Cajal-body specific RNAs*, that localize in the Cajal bodies (nuclear compartments of proliferative and metabolically active cells), where they catalize the methylation and the pseudouridylation of small RNAs.

Additionally, a large number of other snoRNAs displaying noncanonical characteristics and expression patterns have been described.

There are already evidences showing a connection between snoRNAs and cancer. In particular we can refer to an interesting study performed by the group of Chu centered on the role of a specific snoRNA, ACA11. They demonstrated that it is overexpressed in bladder, colon, and esophageal cancers. This altered expression is correlated to the chromosomal translocation (4; 14) that is typical of multiple myelomas and other malignancies. *ACA11* is located in one intron of the gene codifying for the histone methyltransferase WHSC1. The overexpression of *ACA11* promotes resistance to doxorubicin, cellular response to stress, and the proliferation of multiple myeloma cells. Thus, *ACA11* seems to play a role as oncogene and it could be of great interest to deep its potential clinical role (Chu et al., 2012). Other snoRNAs, such as SNORD3A and SNORD60, are involved in cellular stress response, even if a direct connection with cancer has not been defined till now (Cohen et al., 2013; Brandis et al., 2013).

Piwi RNAs (piRNAs)

Piwi RNAs (piRNAs) represent a quite new class of small ncRNAs, 26-31 nucleotide long. The name is due to their association with Piwi protein. Their maturation follows a Dicer-independent mechanism that is different from those of other classes of small ncRNAs.

In 2006 they were first described in mouse sperm cells. They regulate genome integrity, transposon repression, and epigenetics (Aravin et al., 2006; Grivna et al., 2006). Their role in germline development is essential, and in the last years their involvement in cancerogenesis

is arising (Cheng et al., 2012; Yan et al., 2015). Piwi protein seems to be involved in CRC development (Zeng et al., 2011) and, furthermore, different piRNAs have been associated with specific kinds of cancer. Specifically, piR-651 seems to be involved in gastric cancer, piR-823 in gastric cancer and in multiple myeloma (Yan et al., 2015; Cheng et al., 2012), and piR-Hep1 in hepatocarcinoma (Law et al., 2013). Interestingly, these piRNAs regulate different pathways involved in carcinogenesis. In particular, piRNA-823 mediates DNA methylation and angiogenesis, whereas piR-Hep1 is involved in cellular proliferation, motility, and invasion.

Very recently, piRNAs have reached the interest of PGx studies, which were able to demonstrate the association between SNPs in genes codifying for piRNAs and cancer risk (Chu et al., 2015).

These results highlight the great potentiality of the analysis of these ncRNAs, that in future probably will give interesting insights not only in cellular physiology but maybe will have an impact also in patients' treatment.

Vault RNAs

Vault-RNAs (vRNAs) can be detected in cells in the free form or, in a minor extent, as part of the largest ribonucleoprotein complexes, the vaults. They are eukariotic barrel-shaped particles with a mass of 13 MDa and overall dimensions of 400 × 400 × 700 Å (Kedersha et al., 1991). These RNAs are quite abundant in cells and show a high evolutionary conservation, prompting us to hypothesize their key role in cell physiology (van Zon et al., 2003).

In humans, four vRNA genes were identified, *vRNA1-1, vRNA1-2, vRNA1-3,* and *vRNA2-1*. These genes are located in chromosome 5. On chromosome X the *vRNA3-1* is located, but it seems not to be expressed (van Zon et al., 2001). These sequences show 84% of identity and they have similar secondary structures. They regulate mRNAs through the binding with complementary regions located on 5' or 3' UTR. Moreover, it was demonstrated that vRNAs can also generate miRNAs through a Dicer-dependent mechanism (Izquierdo et al., 1998). Interestingly, different lines of evidence demonstrate the importance of vRNAs in multidrug resistance in cancer cells. Specifically, *vRNA1-1* and *vRNA1-2* directly bind to chemotherapeutic drugs like mitoxantrone, probably promoting the export from cells or at least the interaction with their targets (Gopinath et al., 2005). Moreover, vRNAs seem to be also involved in the mechanism of resistance to cisplatin that promotes apoptosis through the involvement of p53. Intriguingly, in *p53*-wild type HeLa cells the inhibition of *vRNA2-1-5p* promotes apoptosis. A direct interaction among vRNA2-1-5p, p53, and its downstream

regulators like Bax and p14, can thus be suggested (Kong et al., 2015; Li J.-H.et al., 2011; MacLeod et al., 1995).

These data support the importance to deep this field in order to highlight new potential targets in cancer treatment.

A special class of nc-RNAs: miRNAs

miRNAs are small endogenous ncRNAs of about 22 nucleotides defined as "*micromanagers of gene expression*"(Bartel and Chen, 2004).

The number of miRNA sequences is constantly growing: the latest version of the Sanger Institute miRBase (release 21) reports 2588 mature miRNAs sequences in human.

Genes codifying for miRNAs can be located in exons, introns or in intergenic not codifying regions. A lot of miRNA genes, called miRtrons, are located in the introns of their own target genes providing a negative feed loop that controls mRNA translation (Okamura et al., 2007).

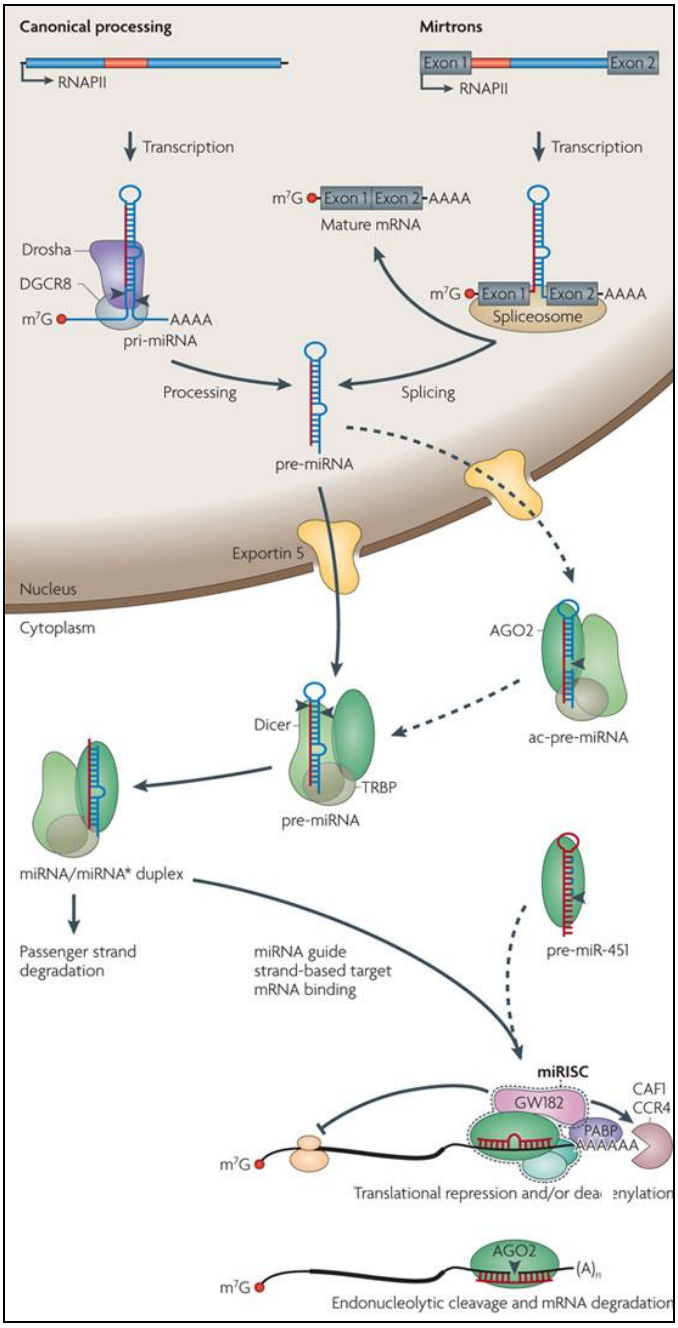

Figure 3: *miRNA biogenesis pathway* (modified from (Krol et al., 2010)). RNAPII: RNA polymerase ii; m7G: 7-methylguanosine cap; PABP: polyADP binding protein; Ago2: argonaute 2; ac-pre-miRNA: Ago2-cleaved precursor miRNA; miRNA*: miRNA passenger strand.

The miRNA biogenesis involves a multitude of proteins located both in the nucleus and in the cytoplasm (*Figure 3*). miRNA genes are transcribed by RNA polymerase II (RNAPII), obtaining the so-called primary transcript (pri-miRNA), that is characterized by a cap and a poly(A) tail (Cai et al., 2004). It presents imperfect hairpins that are the substrate for Drosha and DGCR8 (DiGeorge syndrome critical region gene 8). After this processing, pri-miRNA is converted into premature miRNA (pre-miRNA), a stem-loop structure of 70 nucleotides

with two nucleotides 3' overhang. The pre-miRNA is than exported in the cytoplasm by exportin 5 (XPO5), a RanGTP-dependent double strand RNA (dsRNA)-binding protein.
In the cytoplasm another cleavage takes place directed by Dicer, an endoribonuclease of RNAse III family that forms a complex with TRBP (TAR (HIV-1) RNA binding protein 2), a dsRNA binding protein. After the cleavage, a duplex miRNA is generated, constituted by the guide strand and the passenger strand, known as miRNA*, which is usually degraded. According to the thermodynamic stability, the guide strand is loaded on the RNA-induced silencing complex (RISC), that is formed by various proteins such as Dicer, Ago2 (Argonaute 2), Gemin 3 (Gem-associated protein), Gemin 4, Mov10 (Moloney leukemia virus 10), Imp8 (Importin 8) (Schwarz DS et al, 2003). Once inside the RISC, the mature single strand miRNA binds the 3' untranslated region (3'UTR) of its target mRNA. The complementarity between the miRNA seed region, the sequence spanning from the second to eighth nucleotide from the 5'end of the miRNA, and the target determines the effect on mRNA:
- a perfect matching leads to mRNA degradation, that can be induced by deadenylation, decapping, and exonucleolytic digestion (Rehwinkel et al., 2005; Wu et al., 2006);
- an imperfect pairing leads to inhibition of mRNA translation, that is mainly mediated by Ago: it can prevent cap-recognition by eIF4E (eukaryotic translation initiation factor 4E) or can recruit ribosome inhibitory factors like eIF6 (eukaryotic translation initiation factor 6). As an alternative, mRNA can be also relocated in cytoplasmic foci called p-bodies (processing bodies), which do not contain ribosome (Liu et al., 2005).
Nonetheless, other functions of miRNAs have also been described:
- *translational upregulation of target mRNAs*: this happens under certain conditions, such as quiescence, possibly by the recruitment of mRNAs stabilizing factors (Vasudevan et al., 2007);
- *transcriptional control*: XPO1 and Imp8 regulate the nuclear-cytoplasmic shuttling of miRNAs processed in the cytoplasm and complexed with AGO proteins (Castanotto et al., 2009)(Weinmann et al., 2009). Once in the nucleus, miRNAs can promote the formation of a complex between AGO proteins and transcriptional repressors, inducing the formation of heterochromatin at promoters of target genes. Nevertheless, miRNAs can also activate the transcription of target genes. For instance, miR-373 promotes the transcription of *E-Cadherin* (*CDH1*) in prostate cancer cells (Place et al., 2008). The precise molecular mechanism underlying transcriptional activation is not still elucidated, even if it seems that components of miR-machinery and other epigenetic events are involved (Gagnon and Corey, 2012; Li et al., 2006; Huang and Li, 2012);

- *response to DNA damage*: as well described in the review of Zhang, almost every aspect of cellular response to DNA damage requires miRNAs, that actively participate in DNA damage sensing, damage signals transduction, DNA repair, regulation of cell cycle checkpoints and apoptosis (Zhang et al., 2015).

A higher level of complexity of miRNA activity derives from the presence of target sequences of miRNAs also on long ncRNAs, contributing to the formation of a complex cellular network among different species of transcripts. In particular, long ncRNAs can act both as "miRNAs sponges", preventing them to interact with the target mRNAs, and as direct targets of miRNAs. For instance, the already mentioned long ncRNA HOTAIR is specifically targeted by miR-141, that suppresses HOTAIR expression and function (Chiyomaru et al., 2014).

It is estimated that more than 30% of the human genes are post-transcriptionally regulated by miRNAs, highlighting their capability to deeply influence the majority of the pathways (MacFarlane and Murphy, 2010). The interest about miRNAs is also due to the role they play in pathological conditions, like cancer, fact that leads many groups to study their potential role as biomarkers. Since their discovery, more than 20,400 publications have addressed the association between miRNAs and cancer, including almost 3,700 reviews (PubMed-analysis done on the 16th March 2016).

Rational for a role of miRNAs in treatment outcome prediction

It is well known that miRNAs impact the expression level of several drug-related genes. According to their function, drug metabolizing enzymes can be broadly classified in phase I enzymes, that render drugs more soluble, and phase II enzymes, that transfer moieties from a cofactor to a substrate. Different pharmacogenetic studies about these factors have been published till now, even if only few of them reached a real clinical impact. New and not still well explored mechanisms, such as RNA interference, can give interesting insights in drug metabolism control and, maybe, could give new suggestions in treatment personalization.

The phase I enzymes catalyze three different reactions: oxidation (cytochrome P450-CYP450-, dihydropyrimidine dehydrogenase-DPYD-, alcohol dehydrogenase, …), reduction (nicotinamide adenine dinucleotide phosphate (NADPH)), hydrolysis (epoxide hydrolase, esterases, and amidases). Among them, three subfamilies of CYPs, including CYP1, CYP2, and CYP3, contribute to the oxidative metabolism of more than 90% of clinically administered drugs (Bluth and Li, 2011).

Many efforts have been spent to clarify the role of CYP1B1 isoform in cancer. This interest is due to its overexpression in many types of cancer and its involvment in premalignant progression (Rochat et al., 2001; Luby et al., 2004), data that sustain its potential role as

drug target. Moreover, a clear association between this enzyme and docetaxel was already demonstrated (McFadyen et al., 2001) and very comprehensive reviews about the PGx of this enzyme are available (Sissung, 2006; Li C. et al., 2015). In regards to miRNAs, a strict correlation among CYP1B1, miRNAs, and inflammation has been demonstrated in CRC (Patel et al., 2014). Interleukin-6 (IL6), an inflammatory cytokine highly expressed in CRC microenvironment and with a known prognostic value for CRC (Guthrie et al., 2013), can alter miRNA expression profiles (Lukowski et al., 2015). The interesting study of Patel demonstrated IL6 downregulates *in vitro* miR-27b via DNA methylation, that in turn results in CYP1B1 overexpression and in altered drug metabolism (Patel et al., 2014).

One of the most studied CYP3 enzyme is represented by CYP3A4 isoform. It is involved in the phase I metabolism of many classes of oncologic drugs like hormones (exemestane, tamoxifen), tyrosine kinase inhibitors (erlotinib, gefitinib, imatinib), taxanes (docetaxel, paclitaxel), and alkilating agents (cyclophosphamide). It presents only marginal genetic regulation but its expression level is quite variable among individuals. The 3'UTR of CYP3A4 is targeted by miR-27b. This miRNA is associated with CYP3A4 down-regulation and with *in vitro* chemoresistance to cyclophosphamide (Pan et al., 2009). Moreover, miR-27a regulates also the expression of CYP3A4-transcription factors, retinoid X receptor alpha (RXRα) and vitamin D receptor (VDR), highlighting a complex network in the miRNAs cellular control of CYP3A4 transcription (Ji et al., 2009).

Finally, we can mention another study, published by Lamba, displaying a more comprehensive analysis of the network underlying miRNAs and different isoforms of CYP450 in hepatic tissue. Through a bioinformatic approach and expression analysis, the key role of miR34a and miR148a in regulating CYP3A4 and CYP2C19 came into the light. Considering that these miRNAs are oncosuppressors and have an altered expression in hepatic cancer, the obtained results can be of particular interest in clinics (Lamba et al., 2014).

According to their function and to the clinical value of their genetic variants, the most important phase II enzymes in the oncology setting are represented by uridine diphosphate glucuronosyltransferase (UGT), sulfotransferase (SULT), glutathione S-transferases (GST), N-acetyltransferase (NAT), and thiopurine methyltransferase (TPMT).

Pharmacogenetic analysis of UGT1A1 isoform exerts a considerable impact on ironotecan response. Its importance has led to the introduction of PGx recommendation in the drug label (*Table 1*). This isoform exhibits high interindividual differences in gene expression that can be also explained with epigenetic mechanisms like miRNAs. The determination of such mechanisms could be of great interest in defining tools to help clinicians in treatment

personalization. The group of Dulzen demonstrated *in vitro* that miR-491-3p negatively regulates UGT1A1 expression and thus the metabolism of raloxifene, an oral selective estrogen receptor modulator prescribed to prevent osteoporosis in postmenopausal women. At the best of our knowledge, no data about the effect of this miRNA on UGT1A1 and chemotherapeutic drugs are available, even if the same miRNA gave interesting insights in glioblastoma progression (Li X et al., 2015), opening maybe the way to new scenarios that can in the future be translated into clinics.

Concerning drug activity and patients' response to treatment, it is necessary to briefly introduce also another class of genes, represented by uptake and efflux transporters that regulate cellular intake and uptake of administered drugs. These genes can be roughly classified into two families, the ATP-binding cassette (ABC) and the solute carrier (SLC) factors.

The ABC transporters extrude many substrates using ATP as an energy source. Among them, ABCB1, ABCC1, ABCC2, and ABCG2 play a key role in the oncology field, affecting drug disposition and response (Li and Bluth, 2011).

ABCG2 regulates the cellular disposition of several drugs like anthracyclines, irinotecan, imatinib, and tamoxifen. The 3'UTR displays different target sites for miRNAs, like miR-519c and miR-328. These miRNAs resulte ABCG2 expression in breast cancer cells and can interfere with intracellular drug accumulation (Li X et al., 2011; Pan et al., 2009). In addition, miR-328 expression is down-regulated in side populations of CRC cells characterized by high ABCG2 levels (Xu et al., 2012).

The correlation between miRNAs and ABCB9, a transporter of doxorubicin, was analyzed in an *in vitro* model of lung cancer by the group of Dong (Dong et al., 2014). This study clarified the key role of miR31 in regulating ABCB9. Notably, this miRNA was also significantly associated with the resistance to 5-fluorouracil (5-FU) in early stage colon cancer (Wang et al., 2010), even if, to the best of our knowledge, the mechanism at the basis of this observation has not been elucidated till now.

SLC family includes 360 genes classified in different subfamilies, like organic anion transporting polypeptides (OATP), organic cation transporter (OCT), and organic anion transporter (OAT). Many literature data exist about the interaction between this class of genes and miRNAs. For instance, a significant association between a SNP located in a miRNA-binding site in *SLC19A1* (rs1051269) and methotrexate (MTX) serum concentration in children with acute lymphoblastic leukemia was found. This could be explained suggesting a possible involvement of miRNAs in gene regulation (Wang et al., 2014).

To conclude, the regulation of drug metabolizing enzymes and of genes involved in drug transport exerted by miRNAs justifies the importance to study them as potential tools in treatment personalization. Moreover, very interesting data have been highlighted by the first PGx studies addressing miRNA issue that will be the topic of next paragraph.

Pharmacogenetics of miRNAs: the miRNAs signature

Genetic analysis on miRNAs can be performed evaluating their expression or SNPs located in miRNA genes, in their target regions, or in factors involved in their maturation and activity.

miRNAs expression levels can be easily determined due to their short length that confers them high stability. Moreover, the feasibility of this analysis can also be ascribed to the complex they often form with proteins that protect them from degradation exerted by nuclease. Thus, many groups have analyzed them in different kinds of biological samples like blood, saliva (Momen-Heravi et al., 2014), urine (Sapre et al., 2014), and milk (Munch et al., 2013). The high molecular stability of miRNAs was demonstrated also in formalin-fixed and paraffin-embedded (FFPE) tissues. Actually miRNAs' profiling in these samples and in fresh tissues generated similar patterns, thus opening the way to the big potentialities related to the use of archival specimens for miRNA expression studies (Hall et al., 2012).

An alteration of miRNA levels in cancer tissues can be caused by gene amplification, mutations, polymorphisms, epigenetic alterations in miRNA encoding genes, or abnormalities in their maturation process (Zhang et al., 2006; Calin et al., 2005; Thomson et al., 2006). The specific pattern of miRNAs levels characterizing each tumour is frequently labelled as "miRNA signature". Its definition may be of help to clinicians (Cappuzzo et al., 2014).

The different expression of miRNAs may play various biological roles. miRNAs have a pivotal role in determination and/or maintenance of lineage during development, so the reduced differentiation typical of cancer cells could be ascribed to the anomalous pattern of miRNAs (Lu et al., 2005). Moreover, they can act as oncogenes or oncosuppressors, so their altered levels can impact cellular viability and proliferation (Kjersem et al., 2012). This was demonstrated for CRC. The group of Bartley analyzed a microarray of 866 miRNAs in 69 matched specimens of a microsatellite-stable colorectal adenocarcinoma arising in a contiguous precursor adenoma, and nonneoplastic mucosa of 21 CRC patients. This study highlighted numerous miRNAs differently expressed during progression from adenoma to adenocarcinoma, data supporting a potential active role of miRNAs in cancerogenesis (Bartley et al., 2011).

Great clinical impact can be exerted by miRNAs which expression is altered by therapy. This topic was already well discussed elsewhere (Dreussi et al., 2012), reporting literature evidences coming from *in vitro* and *in vivo* studies about miRNAs involved in response to therapies administered to CRC patients. During last year, the involvement of miRNAs in drug response was corroborated by further studies, such as that performed on primary tumours from patients with metastatic CRC treated with first-line capecitabine monotherapy within the CAIRO trial of the Dutch Colorectal Cancer Group. Low miR-143 expression is significantly associated with longer patient free survival and was suggested as independent biomarker of response to capecitabine (Simmer et al., 2015).

All these evidences could pave the way to larger epidemiologic studies and clinical applications for tumour specimens as source of useful predictive biomarkers that can help clinicians in treatment personalization.

Pharmacogenetics of miRNAs: SNPs on miRNA-related genes

The activity of miRNAs can be influenced by SNPs located in genes codifying for miRNAs, their target sequences and the factors involved in their maturation. The interaction between miRNAs and mRNAs is deeply influenced by the genetic sequences of the regions of recognition both in mRNA and miRNAs. We referred to the already mentioned review for a better clarification of this topic, especially concerning CRC (Dreussi et al., 2012). Here, we will summarize the most important findings published in the last year about that.

Sequences targeted by miRNAs are mainly located in the 3'UTR of mRNAs, even if there are some evidences supporting their existence also in other regions like the 5'UTR and the coding sequence (Almeida et al., 2012). Thus, SNPs located in miRNA target regions could exert an impact on miRNA activity. A pivotal gene in cancer progression and treatment response is *K-Ras*. Its key role was well elucidated in various malignancies as colorectal, lung, and pancreatic cancer. Moreover, its mutational status in cancer tissue predicts strongly the resistance to anti-EGFR therapies, rendering the genetic analysis mandatory. Nonetheless, only about half of the *K-Ras* wild-type CRC patients benefit of these drugs. This supports the need to completely understand gene expression mechanisms, like the regulation exerted by miRNAs. In regards to this, great efforts have been made to clarify the role of a SNP in K-Ras 3'UTR called let-7 complementary site (LCS6-rs61764370), that alters the binding with let7, a tumour suppressor miRNA. The variant allele of this SNP was associated with an *in vitro* increased expression of K-Ras (Chin et al., 2008). The potential clinical impact of LCS6 as predictive biomarker has given till now conflicting results. However, comparing the papers addressing this issue, the lack of homogeneity arises both in the study design and in the clinical parameters that can explain the lack of reproducibility.

The effect of this SNP was also tested on patients enrolled in clinical trials. In particular, it was analyzed on a group of LARC patients enrolled in the EXPERT-C trial, a randomized phase II study aimed at defining the role of the addition of cetuximab before CRT in the neoadjuvant setting. The genetic analysis was conducted on 155 FFPE cancer tissue samples. The presence of the G allele is significantly associated with complete response after neoadjuvant treatment, showing also a trend for better 5-year progression free survival (PFS)(Sclafani et al., 2015). LCS6 was also investigated in advanced non-small cell lung cancer patients enrolled within the TAILOR trial, which compared erlotinib and docetaxel as second line treatment. In the docetaxel arm, patients with the genotype TT/TG experienced a longer PFS compared with GG patients (Ganzinelli et al., 2015). These preliminary results claim for further studied to better elucidate the predictive and the prognostic value of LCS6.
Another class of SNPs of pharmacogenetic interest is represented by those located in miRNAs. Nevertheless, miRNAs codifying genes highlighted a high grade of conservation with very rare sequence alterations, probably due to their major role in the control of cellular gene expression. This supports the hypothesis that considers the high biological impact of SNPs located in these genes. The genetic analyses are usually not strictly directed on mature sequences of miRNAs but also on their precursors. For instance the group of Tahara found a significant association between rs3746444 located on *premiR-499* and prognosis of advanced gastric cancer patients treated with chemotherapy, in terms of OS and PFS (Tahara et al., 2014). To conclude, we consider appropriate also to cite two very recent works showing the significant prognostic role of two SNPs located in two different miRNA genes, *miR-608*-rs4919510 and *miR-492*-rs2289030 in hepatocellular carcinoma patients who have undergone surgery (Yu et al., 2016; Ma et al., 2016). These results are encouraging considering the high aggressiveness of this malignancy and the critical need of robust biomarkers for better patients' management.
Another interesting and still not well explored field is the PGx of factors involved in miRNAs' maturation. For the sake of simplicity, a brief overview of data regarding Dicer, one of the main key players of miRNA maturation, is here presented. Dicer expression levels were determined in different tumours, and the results were controversial: in prostate, oesophageal, and oral cancer it was upregulated (Chiosea et al., 2006; Sugito et al., 2006; Jakymiw et al., 2010), whereas its downregulation was detected in breast, ovarian, and advanced lung cancer (Grelier et al., 2009; Merritt et al., 2008; Karube et al., 2005). An altered expression of Dicer probably contributes to the specific patterns of miRNAs expressed in cancer tissues. Notably, some evidences showed a significant association between the Dicer mRNA levels and the clinical outcome of patients. The first evidence

supporting the predictive value of Dicer expression in cancer tissues of patients affected by CRC was provided by Vincenzi and colleagues (Vincenzi et al., 2013). In this study, Dicer mRNA levels were assessed in 116 FFPE tissues of advanced CRC patients treated with bevacizumab. Low expression of Dicer was significantly associated with PFS, OS, and better response to bevacizumab. This can be explained considering the impact of Dicer on miRNA expression that ultimately affects cell physiology. Moreover, Dicer is involved in other patways, such as angiogenesis, which could be related with the reported results too (Kuehbacher et al., 2007; Suárez et al., 2007).

All these interesting data and the scarcity of results available till now support the great potential of the genetic analysis of SNPs impacting miRNAs activity and maturation.

In the next paragraph, we will present a new field of study of PGx, immunogenetics. In the last years it is increasingly drawing the scientific community attention, thanks also to the interesting results coming from immunotherapies in oncology.

IMMUNITY AND CANCER

The major role of immune system is to preserve the tissue homeostasis, protecting the organism from the attack of pathogens and eliminating damaged cells. A possible link between cancer and immune system was hypothesized for the first time in 1863 by Rudolf Virchow, who noticed the presence of leukocytes in neoplastic tissue (Balkwill and Mantovani, 2001). The primary role exerted by immune system in cancerogenesis arises overwhelmingly considering the incidence of cancer in patients affected by immune deficiency syndromes. It was indeed extensively observed that people infected with HIV are thousand times more at risk to develop Kaposi sarcoma, 70 times more at risk for non-Hodgkin lymphoma, 5 times more at risk to develop cervical cancer, and they show also a higher risk to be diagnosed with several other malignancies such as anal, liver, and lung cancer (Grulich et al., 2007). Evidence supporting the key role of immune system in cancer establishment is represented by people undergoing iatrogenic immunosuppression following organ transplantation, who are at higher risk to develop various malignancies compared to general population. For instance, patients who had received heart transplants have a median 2.7 fold increased risk of cancer development compared to the general population. The risk varies according to cancer type (Jiang et al., 2010).

All the well-known hallmarks of cancer described by Hanahan and Weinberg are somehow influenced by the immune system (Hanahan and Weinberg, 2011). In particular, there are at least three activities of the immune system that can be included under the umbrella of the mechanisms protecting the organism from cancer onset:

- it protects the body from the infection of human tumour viruses such as hepatitis B virus (responsible for liver cancer), papilloma virus (responsible for cervical and other anogenital cancers), Epstein-Barr virus (responsible for Burkitt's lymphoma and nasopharyngeal carcinoma), human T-cell lymphotropic virus (responsible for adult T-cell leukemia) (Liao, 2006). These infections are estimated to cause the 7.0% of cancers in Europe, percentage that increases reaching the 16.1% considering cancer burden all over the world (De Martel et al., 2012);
- after the elimination of pathogens, the immune system promotes the resolution of inflammation, a condition that favors cancer onset. This arises clearly observing the higher incidence of cancers in patients affected by chronic inflammation. These conditions are characterized by a production of reactive oxygen and nitrogen compounds that damage nucleic acids, thus resulting in cell damage and possible tumorigenesis;
- it can also identify and destroy nascent tumour cells because they express anomalous antigens on their surface, with a mechanism defined immunosurveillance.

Nonetheless, the immune system can be overwhelmed by cancer cells that are also able to take advantage of it. This can be explained referring to the so called *immunoediting hypothesis*, which is a dynamic process that includes three stages: elimination, equilibrium, and escape (*Figure 4*).

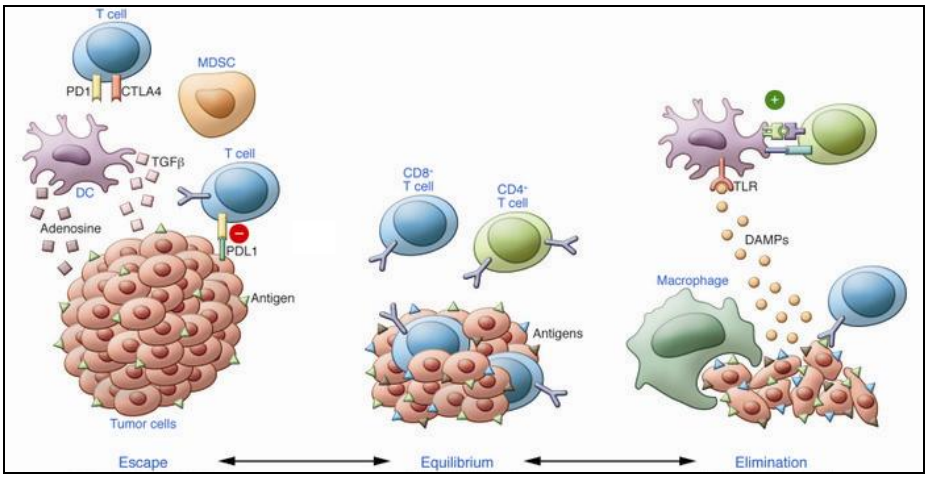

Figure 4: *Immunoediting*: the involvement of immune system in cancerogenesis can be depicted with the immunoediting hypothesis, which specifically identifies three different stages: escape, equilibrium, and elimination (modified from (Kalbasi et al., 2013)).

The *elimination phase* is the new way to conceive the already described mechanism of immunosurveillance that leads to the removal of mutated cells that expose on their surface immunogenic epitopes. Moreover, in this stage stress and necrosis represent immunostimolatory signals helping immune system activity. In the *equilibrium phase*, tumour cells are dormant and not still clinically detectable also for years. This condition is

characterized by a dynamic balance between immune system and tumour. If cells acquire new mutations or if the activity of the immune system is somehow altered, tumour starts to grow entering the *escape phase*.

Cancer immunoediting, immune system and inflammation are strictly correlated phenomena. Chronic inflammation can support cancer initiation (generating genotoxic stress), cancer promotion (supporting cell proliferation), and cancer progression (affecting angiogenesis, epithelial to mesenchimal transition, and metastatization). As well described in the review of Grivennikov (Grivennikov et al., 2010) there are different types of factors that can induce tumour-promoting inflammation:

- infections (such as the aforementioned hepatitis B virus and Helicobacter pylori);
- particulate material and other components of tobacco smoke;
- obesity, which is associated with an higher risk of liver cancer;
- autoimmune disease, like celiac disease, that is related with an augmented risk of liver cancer.

Intriguingly, in cancer samples there are always inflammatory infiltrates. This can be explained by the capability of cancer cells to exploit immune response in different ways. It was demonstrated for instance that myeloid cells preferentially differentiate into myeloid derived suppressor cells or M2 macrophages that produce immunosuppressive and proangiogenic factors inside the tumour microenvironment, thus promoting cancerogenesis (Gabrilovich and Nagaraj, 2009; Colotta et al., 2009). Moreover, malignant cells secrete proinflammatory factors in response to oncogene activation (e.g. *Ras*, *Myc*) and cell senescence induced by DNA damage, typically occurring in cancer cells. Inflammation can also be responsible for epigenetic changes, including DNA methylation and altered miRNA expression that can consequently silence oncosuppressors. All these phenomena contribute to the onset of the tumour microenvironment that is crucial for other mechanisms like angiogenesis and metastatization, which in turn have an impact on tumour development and on treatment response.

The complex interplay between inflammation and cancer is better summarized in *Figure 5*.

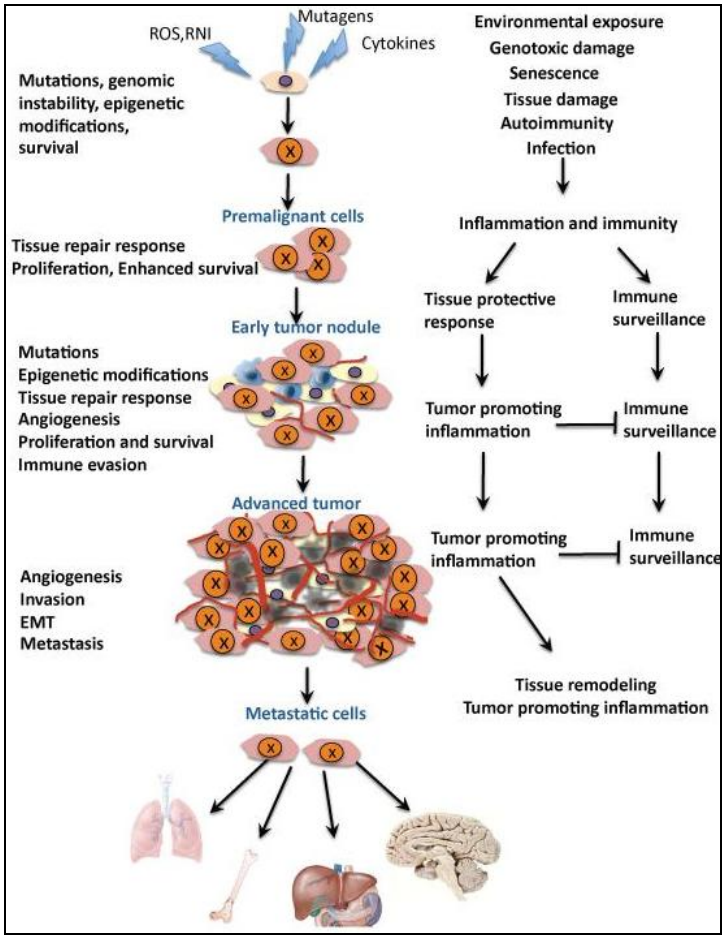

Figure 5: *The different roles of inflammation in cancer*: inflammation acts at all stages of tumorigenesis. Mutated cells are marked with "x". Yellow - stromal cells, brown - malignant cells, red - blood vessels, blue - immune and inflammatory cells. Epithelial-mesenchymal transition, EMT; reactive oxygen species, ROS; reactive nitrogen intermediates, RNI (modified from (Grivennikov et al., 2010)).

Immunogenetics: the new watershed in PGx?

It is clear that immune system acts as a double-edged sword in cancer battle, and the deeply definition of such mechanisms will for sure lead to the development of new strategies to harness the power of immunity to defeat cancer. This is the basis of immunotherapies. The definition of immune system role in cancer can be faced from different points of view, bearing in mind first of all that its complex activity is orchestrated by different kinds of cells and interleukins, as deeply described in the review of Vesely (Vesely MD et al, 2011). One possible way to go in deepening this matter is to understand the potential role of SNPs in genes codifying for factors involved in immune system activity, as well as the role of their altered expression. Some groups have already adopted this approach leading to the definition of a new field of study, immunogenetics.

We should underline that much is already known about immunogenetics and virus-associated tumours and about immunogenetics and cancer associated with conditions

affecting immune response, whereas much is to do regarding malignancies not caused by infections.

Regarding the last scenario, immunogenetics may be useful to identify predictive biomarkers. The role of immune system in therapy can be explained considering that chemo- and radiotherapy promote the necrosis of malignant cell. This mechanism results in the loss of cell membrane integrity and in the release in the extracellular space of proteins and other internal cellular components, factors with a high proinflammatory activity (*Figure 6*). In this case, an opposite effect can be exerted by immune response. On one hand, the activation of immune system can promote antigen presentation, thus inducing tumour eradication. On the other hand, the macrophages infiltrating the bulk can be activated, promoting the release of cytokines with oncogenic activity.

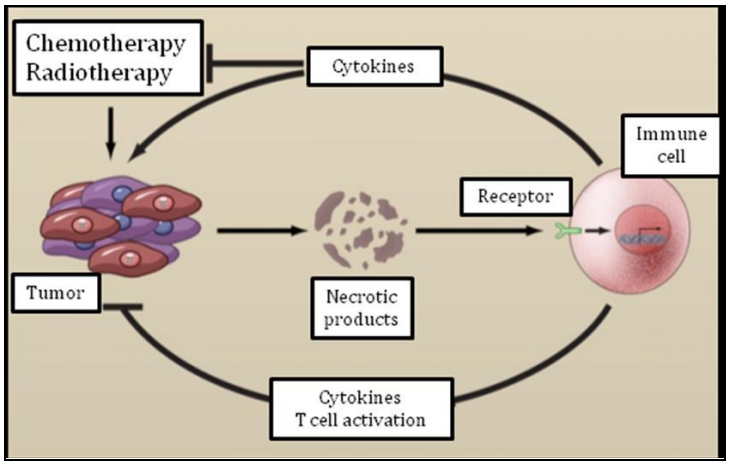

Figure 6: *Immune system role in treatment response*: Therapy-induced inflammation and immune response to chemo- and radiotherapy (modified from (Grivennikov et al., 2010)).

It is quite conceivable to hypothesize that the major role of immune response in therapy response is played by two families of factors: those infiltrating tumour bulk and those located at circulating level that can exert a systemic affect. Among these proteins, the most important mention inflammatory mediators are cytokines, transcription factors activated by Nf-kB and STAT3, and molecules regulating angiogenesis and invasiveness.

One commonly used approach applied for the immunogenetic analysis is represented by the study of biomarkers expression. Among them, TGF-α upregulation is associated with tumour downstaging in rectal cancer patients treated with cetuximab and capecitabine. Moreover, an altered expression of many immune factors such as immunoglobulin M (IgM), interleukin 4 (IL4), and tumour necrosis factor-β (TNF-β) could predict tumout recurrence (Debucquoy et al., 2009).

Another field of study of immunogenetics is represented by the analysis of SNPs located on inflammatory genes. There are already evidences about the association of these genetic variants and cancer risk (Landi et al., 2003; Landi et al., 2007; Bai et al., 2013). The role of SNPs in immune-related genes as predictive or prognostic biomarkers has been also investigated, even if the scarcity of data claims the necessity to further deep this field of study.

The group of Schoenfel found a significant association between rs45757998 located in *ribonuclease L* and the outcome of 434 early-stage prostate cancer treated with radiotherapy. Intriguingly, the variant allele of this SNP was associated with augmented serum levels of C-reactive protein and IL6, thus having an impact on the global inflammatory state of the patients (Schoenfeld et al., 2013). Expression of IL6 plays a prognostic value in CRC patients: it is indeed shown that high tissue levels is associated with poor survival (Chung et al., 2006). IL6 can be ascribed among the key cytokine networks in CRC, due to the role it exerts on cancer progression, migration and angiogenesis (West et al., 2015). Its pivotal activity in this scenario led some authors to investigate the role of SNPs located in this gene. Interestingly, the analysis of 4 SNPs in *IL6* (rs1800797, rs1800796, rs1800795, and rs2069849) has been proposed as a useful tool for clinicians to predict early events in estrogen receptor negative breast tumours (Markkula et al., 2014). Moreover, IL6 serum levels have been associated with breast cancer prognosis. We can hypothesize that IL6, despite cancer site, can be used as prognostic biomarker, even if this idea should be ascertained with further analysis.

Another important cytokine is represented by IL17. This molecule is involved in the pathogenesis of various autoimmune diseases and it is also considered as an important mediator of cancer-associated inflammation and of angiogenesis. This observation led some authors to suggest the use of IL17 as prognostic biomarker in CRC patients and, moreover, to consider it as a new possible drug target. The main isoforms of this cytokine are IL17A and IL17F, which are modulated in an opposite way. It is indeed observed that IL17A levels are upregulated in blood and cancer tissue, whereas those of IL17F are downregulated. *IL17F*-rs763780 and *IL17A*-rs2275913 were analyzed in a group 102 CRC patients by the group of Omrane: the SNP of IL17A could affect response to CRT, while the one of IL17F can be used as an independent prognostic biomarker (Omrane et al., 2015).

Finally, we think it is noteworthy to bring to the attention also some studies analyzing SNPs in vascular endothelial growth factor (VEGF), which mediates both angiogenesis and inflammation. A very comprehensive meta-analysis identified the crucial role of rs2010963 (*VEGF* 405C>G) in the prognosis of different kind of cancers (Eng et al., 2012). In literature

there is plenty of other papers investigating the role of SNPs in this gene or in factors related with angiogenesis and tumour response, also recently published (Sohn et al., 2014).
To conclude, it is notably that the French group headed by Galon have prompted the use of the so called "*Immunoscore*". They analyzed a large cohort of CRC patients and demonstrated that there is an association between the immune milieu in which cancer arises and patients' prognosis. The immunoscore considers type, density of two lymphocyte populations (CD3+ and CD8+), and location of immune cells in cancer specimen. According to these parameters, this score ranges from 0 to 4, *i.e.* from low to high density of infiltrating immune cells (Galon et al., 2012). Its prognostic value was ascertained in different studies, which demonstrated even a better prognostic power then TNM staging (Galon et al., 2013).
Despite the wave of optimism and hope generated by these immune signatures, in the last years a general stall of the translation of these concepts into the clinical practice has taken place. Nonetheless, the great potentialities correlated with a full understanding of this mechanism are undeniable.

3.2.2 EVOLUTION OF PHARMACOGENETICS : NEW TOOLS

Knowledge in genetics has made giant steps, moving from the first studies that determined the definition of Mendel laws reaching the possibility to analyze the entire human genome. As aforementioned described, great advances have been made also in defining the so called "junk DNA", which hosts key regulatory regions such as those codifying for ncRNAs. In addition, in the last decades, the translation of PGx into clinical practice has started, leading to an update of some drug labelling. However, much is still to do. For instance, the potential clinical validity of SNPs located in non codifying regions and in immune-related genes has to be definitely clarified.
These theoretical advances have been possible thanks to the tremendous technological advances in genetics and biology analysis methods occurred in the last decades. These improvements gave rise to big genetic projects, like the *Human Genome Project* and the *1000 Genomes Project*.
Great efforts have been spent for the Human Genome Project that was officially launched in 1th October 1990 and was finished in April 2003. This was coordinated by the National Center for Human Genome Research, now evolved in the National Human Genome Research Institute (NHGRI), which still continues the implementation of genetic analysis, fostering different branches of research (www.genome.gov). It is amazing to read that some of the main priorities of NHGRI research include "*understanding how the human genome works; establishing the role of genomic variants in health and disease; using genomic information to*

advance medical care and human health; and addressing the societal impact of genomic advances", some problems that, even if obviously in a more simplistic way, are somehow addressed in this thesis.

Another very important project is the 1000 Genomes Project, which is cataloguing sequence variants in the human genome. It is the result of an international collaboration with the aim to produce an extensive public catalogue of human genetic variation, including SNPs and structural variants, and their haplotype contexts. This finally can support genome-wide association studies and other medical research studies. The objective of this project is to analyze the genomes of about 2,500 unidentified people from about 25 populations around the world using next-generation sequencing technologies. The results of this study are freely and publicly accessible to researchers worldwide.

These and other genetic projects were possible thanks to the development of sequencing methods and genotyping technologies. All these novelties have led to fuelling the doubt about the possibility to efficiently understand the biological and clinical impact of this big amount of data . "*Biomedical research and the practice of medicine are reaching an inflection point: the capacity for description and for collecting data, is expanding dramatically, but the efficiency of compiling, organizing, and manipulating these data-and extracting true understanding from them-has not kept peace*" (Towards precision medicine, 2011).

In this thesis, we had the possibility to obtain genetic data from a medium-throughput technology. The analyses were indeed performed with *Illumina BeadXpress Reader* (Illumina, La Jolla, CA), a dual-color laser scanning system, based on Golden Gate technology. With this kind of tool users can perform multiplex testing ranging from genotyping, gene expression, RNA and protein-based assays, methylation and expression studies. The second key technology of BeadXpress system is represented by the VeraCode ™ microbeads, silica microcylinders (240 μm in length by 28 μm in diameter), that are each inscribed with a unique digital holographic code to unambiguously designate and track the specific analyte or genotype of interest throughout the multiplex reaction. Differently from traditional microarrays, these beads are used in solution, allowing more rapid hybridization times. More technical details about this system are presented in the *Materials and Methods* chapter, here we have given only a brief overview to better ascribe the project of this thesis.

Another step forward in PGx is represented by bioinformatics, a discipline that represents a cutting edge field of modern biology. It is a wide science based on the knowledge coming from computer science, biology, engineering, and mathematical methodologies. The use of bioinformatics tools ensures for instance acceleration in the stage of genes and SNP selection that is commonly based only on literature analysis. This is a tricky and pivotal

phase in PGx studies and its importance is sometimes neglected. Another advantage related with the use of bioinformatics is due to the possibility to refer to tools specifically designed for the pathways of interest. In this thesis we selected genes involved in miRNA maturation through the use of programs specific for this pathway like PolymiRTS and microSNiper. However, it is not always possible to exploit such kind of programs as we experienced for the selection of genes involved in immune system activity in cancer. For this second analysis, another strategy was indeed designed due to the scarcity, at least at our knowledge, of tools specific for this pathway. Moreover, it is to be highlighted that the evolution of bioinformatics is so fast that it is really a challenge to keep up with the continuous updates in computational and biological science. Thus, it is still fundamental to join data achieved through the bioinformatics tools with literature analyses, and this observation led us to design these strategies that join these two approaches, as reported in the *Materials&Methods* chapter. Moreover, even if in a limited extent, we followed an analogous procedure also to interpret the obtained results in the *Discussion*.

3.3 THE CLINICAL PROBLEM: LOCALLY ADVANCED RECTAL CANCER MANAGEMENT

The clinical model analyzed in this thesis is represented by LARC. The estimated incidence of new rectal cancer cases in the United States in 2016 amounts to 39,220 (NIH). The peak incidence of rectal cancer is during the fifth decade of life. Metastases arise most frequently in liver, followed by lung, retroperitoneum, ovary, and peritoneum.

The complexity of rectal cancer treatment strongly depends on the risks of the side effects, such as urinary and sexual dysfunction, faecal urgency and incontinence.

In this context some problems arise. First, one crucial problem is represented by the selection of the best CRT regimen, both in preoperative (neoadjuvant) and in postoperative (adjuvant) phase. According to the American guidelines, patients should undergo surgery in association with adjuvant treatment (NCCN guidelines). In Europe, the standard treatment is represented by neoadjuvant radiotherapy possibly associated with fluoropyrimidines (5-FU or capecitabine) and platinum derivatives (Glimelius et al., 2013). Neoadjuvant treatment aims at reducing the tumour mass leading to a more sparing surgery choice. Response to this treatment is associated with patients' prognosis and local recurrence control. However, in only 8-15% of cases tumour downstaging/downsizing and complete pathological response are achieved (Aschele et al., 2005; Probst et al., 2015). For non responder patients, other therapeutic strategies should be planned, without delaying surgery and also avoiding useless and potentially toxic CRT. According to this, another

crucial problem is represented by the definition of the best system to classify treatment response. One of the most used scale is the tumour regression grade (TRG), defined according to Mandard's criteria adapted from esophageal carcinoma (Mandard et al., 1994). It is a 5-point histopathological scaling based on the level of fibrosis and on the presence of cancer cells in surgical specimen. Specifically, TRG=1 is associated with pathologic complete response, TRG=2 with the presence of residual cancer cells scattered through the fibrosis, TRG=3 with an increase in the number of residual cancer cells, but fibrosis still predominated, TRG=4 with residual cancer cells outgrowing fibrosis, and TRG=5 with absence of regressive changes. This parameter has also a prognostic value. It is indeed associated with recurrence free survival (Trakarnsanga et al., 2014), disease free, metastases free and overall survival (Roy et al., 2012; Vecchio et al., 2005).

Thus, the clinical dilemma of different response to chemo- and radiotherapy is still not overcome and the identification of the underlying mechanisms and predictive biomarkers represents a pivotal need for optimizing cancer patients' treatment. Moreover, no reliable prognostic biomarkers have been introduced into the clinics. Specifically, two big problems are related with LARC:

- a complete response to neoadjuvant treatment is achieved only by a restricted portion of patients (Aschele et al., 2005; Probst et al., 2015). Thus, the identification of patients who have a low probability to respond to therapy can of great interest as they could avoid neoadjuvant treatment, sparing useless toxicities related to CRT and optimizing the time for the surgical intervention;
- the second problem is related to patients' prognosis. In particular, understanding who will relapse can be of help to clinicians. This can be ultimately translated into a more appropriate follow-up and, potentially, have an impact on the choice to treat patients in the adjuvant setting, that still represents an open question (Petrelli et al., 2015).

These two problems are of vital importance to optimize patients' treatment and define the more appropriate follow-up. In this scenario, the identification of genetic biomarkers can really have a great clinical impact and could play a crucial role in rectal cancer patients' management.

4. MATERIALS & METHODS

4.1 PATIENTS' ENROLMENT , TREATMENT , AND CLINICAL DATA COLLECTION

This retrospective study, organized by the CRO-National Cancer Institute of Aviano, Italy, included patients with LARC treated with CRT in neoadjuvant regimen. Between December 1993 and July 2011, a total number of 520 LARC patients were enrolled: 282 patients were enrolled at the CRO-National Cancer Institute of Aviano, Italy and 238 at the General Hospital of Padua, Italy. A complete clinical history and physical examination, colonoscopy, complete blood cell count, chest X-ray, hepatic and trans-rectal ultrasound, pelvic computed tomography scan, and carcino-embryonic antigen were available for all the patients.

The elegibility criteria for this study were the following ones:

- ✓ histologically confirmed diagnosis of primary LARC;
- ✓ confirmed absence of distant metastases;
- ✓ age ≥18 years;
- ✓ performance status (according to World Health Organization) 0–2;
- ✓ normal bone marrow, renal, and liver function;
- ✓ Caucasian ethnicity;
- ✓ kind of treatment: neoadjuvant treatment based on fluoropyrimidines (either 5-FU or capecitabine) in presence or absence of oxaliplatin, with two distinct dosages of RT (either 50.4Gy or 55Gy), followed by a radical surgery.

All procedures were reviewed and approved by the Institutional Review Board and Ethical Committee of both participating institutions. All patients signed a written informed consent for research purposes before entering this study and agreed to provide a peripheral blood sample for germinal DNA analysis. Based on these criteria, 280 patients resulted eligible.

Patients' medical records were examined to collect the following clinical information:

- ✓ age;
- ✓ gender;
- ✓ date of diagnosis;
- ✓ clinical tumour, nodal, and metastasis stage (cT, cN, and cM, respectively);
- ✓ tumour distance from anal margin;
- ✓ neoadjuvant treatment parameters: RT dosage, fluoropyrimidines route of administration, concomitant platinum administration;
- ✓ date of end of neoadjuvant RT;
- ✓ date of surgery and kind of surgical intervention, specifying if intraoperatory RT

(IORT) was performed;

- ✓ pathological response to neoadjuvant treatment defined as tumour regression grade (TRG);
- ✓ post CRT pathologic T stage (ypT);
- ✓ post CRT, if administered;
- ✓ recurrence;
- ✓ date of last follow-up/death.

Patients underwent external beam RT with a 10-18MV linear accelerator. A 3D-CRT was used in all patients. They were treated in prone position with full-bladder. A dedicated up-down table was used for patient immobilisation and small bowel dislocation outside the target volume as previously described (De Paoli et al., 2006). The primary tumour, the mesorectum, the posterior wall of the bladder and prostate/vagina, and the internal iliac nodes represented the clinical target volume (CTV). Two different RT programs were applied, according to clinical trials ongoing in the considered period time: 202 patients were treated with a standard dose of 50.4Gy/28 fractions, whereas 78 with a dose of 55.0Gy/25 fractions. In the first group, a consecutive boost of 50.4Gy/3 fractions to the tumour and mesorectum was given following the CTV dose of 45Gy/25 fractions, for a total dose of 50.4Gy. In the second group, a concomitant boost of 10Gy/10 fractions over five weeks, two times a week (1Gy/fraction, 6 hours interval between the two daily fractions), was delivered to the tumour and mesorectum during the CTV dose of 45Gy fractions, for a total dose of 55Gy. Fluoropyrimidines alone (5-FU 225mg/m ^{2}/day iv continuous infusion for 5 weeks or capecitabine 1650mg/m^2 in two daily oral administrations for 5 weeks) was prescribed with 50.4Gy or 55.0Gy, whereas the capecitabine (1300mg/m ^{2}) was administered with oxaliplatin (130mg/m^2 every 19 days) and concurrently standards RT dose of 50.4Gy.

Radicality of the surgical procedure and tumour-, node- and metastasis staging was reported following the American Joint Committee on Cancer Tumour, Nodes and Metastasis (TNM) classification. The whole residual tumoral area was sampled for histopathological examination and histopathological tumour-staging after radiotherapy (ypT) assessment. Mesorectal surgical margin status and lymph nodal changes were described in the pathological reports. The pathological tumour response to preoperative therapy was adapted from the TRG criteria suggested by Mandard for esophageal carcinoma (Mandard et al., 1994), and assessed as previously described (Cecchin et al., 2011).

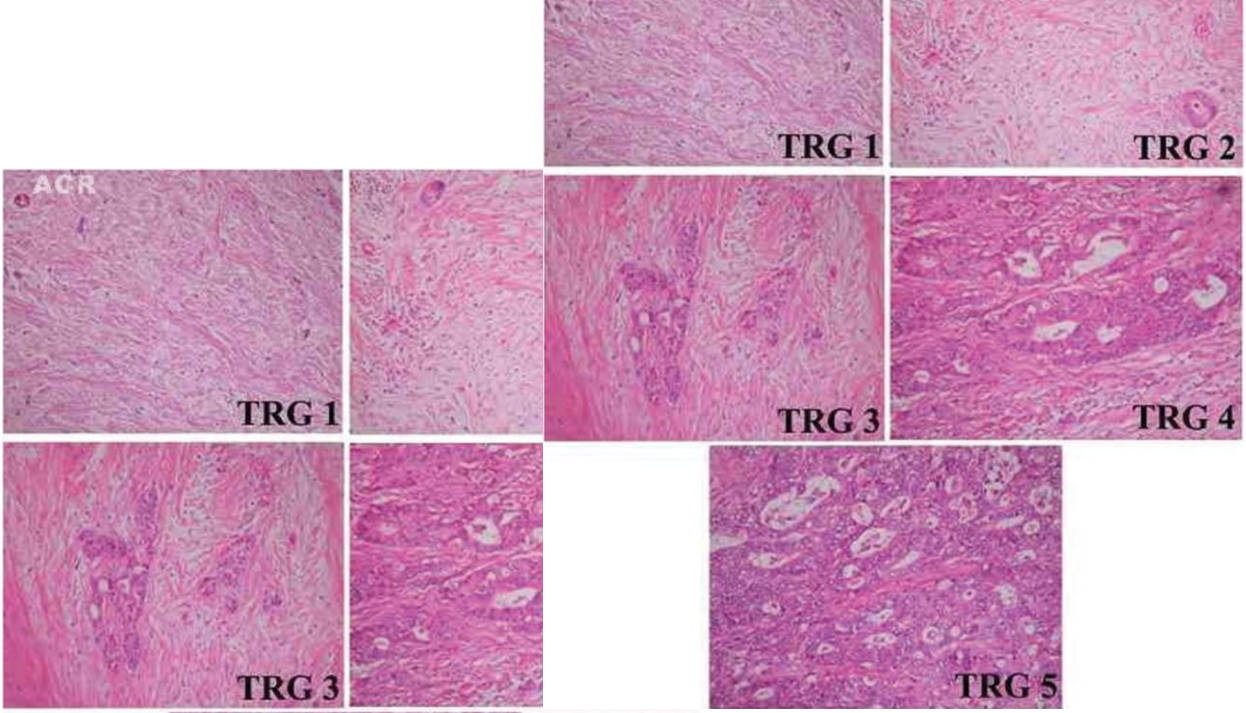

Figure 7: *Tumour Regression Grade:* pathological tumour response to neoadjuvant treatment defined according to the Tumour Regression Grade (TRG) in rectal cancer patients, adopted from the scale proposed by Mandard –Figure adopted from (Santos LG et al, 2012)-.

Specifically, as you can see in the *Figure 7*, TRG=1 is associated with pathologic complete response, TRG=2 with the presence of residual cancer cells scattered through the fibrosis, TRG=3 with an increase in the number of residual cancer cells, but fibrosis still predominated, TRG=4 with residual cancer cells outgrowing fibrosis, and TRG=5 with absence of regressive changes.

All patients were followed-up every three months for the first two years, every six months thereafter up to five years, and then yearly.

Patients' clinical data have been collected by radiologists and oncologists. All personal and clinical data were catalogued in appropriate databases, prepared in accordance with the Privacy Policy, in order to be associated with genetic data.

4.2 BIOINFORMATIC & LITERATURE ANALYSES : SNPS SELECTION

4.2.1 SNPS SELECTION : MIRNA PROJECT

We aimed to select a set of miRNA-related SNPs potentially impacting miRNA maturation and activity.

We referred to Patrocles, a database of miRNAs' genetics information, and to PubMed to identify genes involved in miRNA maturation. In addition, we used an *in silico* approach to predict miRNAs controlling four key factors of this pathway (POLR2A, Drosha, DGCR8, and Dicer), selecting those predicted by at least three algorithms freely available on-line (Patrocles, TargetScanHuman, Microcosm, miRanda, Pictar, PolymiRTS, microSNiper). Thus, we identified 63 genes. The selected genes were submitted to Illumina. The Illumina assay design tool identified a list of 13067 SNPs located in these genes. The predicted final score (which predicts the quality of the assay giving a score ranging from 0 to 1), the designability

(which ensures the capability to design the selected assay giving a score spanning from 0 to 1), and the reported minor allele frequency (MAF) in Caucasian population (HapMap CEU) were considered for the screening. Only SNPs with high predicted final score (≥0.6), optimal designability (=1) and MAF≥5% were selected. Then, tagging-SNPs were chosen with GenomeVariationServer tool according to the following criteria: r^2>80%, data coverage for tagSNP>85%, and data coverage for clustering=70%. For each haploblock, the SNP with the highest predicted final score was selected.

4.2.2 SNPs SELECTION : IMMUNOGENETICS PROJECT

We designed a workflow to select a set of SNPs located in genes involved in immune response analyzable with BeadXpress technology.

Firstly, we performed a literature analysis to identify genes impacting this pathway, referring to “cancer” and “immune system” as mesh terms, prioritizing factors involved in cancer risk and tumour response treatment. We selected 34 genes. We referred to UCSC genome browser to define the genomic coordinates of the these genes. We extended the regions of interest of further 5000 nucleotides up- and downstream of the target gene to include reasonably all the regulatory regions.

The obtained genomic coordinates were then uploaded to HapMap website and Haploview software was used to define blocks of SNPs in strong linkage disequilibrium setting r^2≥0.80 and MAF≥0.05 in Caucasian population.

Polymorphisms were prioritized according to their biological effect and to evidences reported in literature. In particular, their function was determined according to HaploRegv2. The highest priority was given to missense SNPs and SNPs previously associated with cancer or immune system activity. Secondly, SNPs located in regions bound by a transcription factor or other regulatory proteins and SNPs in a promoter or enhancer sequences were selected. The list of SNPs was then uploaded to Illumina website and analyzed with the Illumina assay design tool as aforementioned described. We selected only assays with high final score (≥0.7) and optimal designability (=1).

4.3 MOLECULAR ANALYSES

4.3.1 GENOMIC DNA EXTRACTION

A 3mL whole blood sample was collected from each patient, and stored in freezer at -80°C. Blood specimens were collected in sterile tubes with any anticoagulant agent but heparin was not admitted.

The extraction of genomic DNA from whole blood was performed with the automated extractor BioRobot EZ1 (Qiagen SPA, Milano, Italy). The Card "EZ1 DNA Blood", in association with the Kit "EZ1 DNA Blood Kit 350μL", was used for the extraction of genomic DNA from 350 μl of whole blood obtaining 200μL as final volume. Once introduced the appropriate card and start the program, the BioRobot allows to process six samples simultaneously, without any intervention by the operator.

This technology is based on the use of silica-coated magnetic particles. In one step DNA is isolated from lysates through its binding to the silica surface of the particles in the presence of a chaotropic salt. The particles then are separated from the lysates using a magnet and the DNA is efficiently washed and eluted in a buffer. DNA yields depend on the sample type, number of nucleated cells in the sample, and the protocol used for purification of DNA. For the protocol applied in this thesis, the yield should be approximately of 5-12μg of DNA in 200μL.

The extracted DNA was stored at 2-8°C and its purity determined evaluating the absorbance at different wavelengths and calculating this ratio: (A260 – A320)/(A280 – A320). Pure DNA has an A260/A280 ratio of 1.7–1.9.

4.3.2 METHODOLOGIES FOR SNPS ANALYSES

As already described in the previous sections (4.2.1 and 4.2.2), two different strategies were applied to select the genes and the SNPs of interest. To analyze them, two kind of semi-automated, recently developed genotyping methods have been used: direct sequencing and BeadXpress. Both are based on the polymerase chain reaction (PCR). Thus, in the next paragraphs a brief overview about PCR, Sanger direct sequencing, and BeadXpress are presented.

POLYMERASE CHAIN REACTION

This technique was set up in 1983 by Kary Mullis and allows the *in vitro* production of a large number of copies of a specific DNA sequence, procedure that has a plethora of applications, like sequencing, cloning, mutagenesis procedures, diagnostic tests.

The process is schematically represented in *Figure 8*. During this procedure, the samples are subjected to a series of thermal cycles that are summarized below:

- ✓ an initial period at elevated temperature (5-10 minutes at 94-95°C) that allows the DNA thermal *denaturation*, in order to separate the template strands that act as a mold;
- ✓ a variable number of consecutive *cycles of amplification*, typically 25-40 times. Each cycle consists of three phases:

  1. complete DNA thermal denaturation (94-95°C);

  2. pairing (annealing) of sense and antisense primers with complementary sequences on the DNA template. Primers are single-stranded oligonucleotides complementary to regions located up- and downstream of the sequence of interest. They are mandatory to start the amplification. In this phase the temperature is lowered to values which may vary from 50°C to 65°C;

  3. extension (elongation) of the primers and synthesis of new strands catalyzed by the DNA polymerase. The extension product grows by the formation of a phosphodiester bridge between the 3´hydroxyl group at the growing end of the primer and the 5´phosphate group of the incoming deoxynucleotide. The temperature is usually set at 72°C, that favors the optimal enzyme activity;

- ✓ a *final elongation* step is usually performed at the end of the amplification cycles (10 minutes at 72°C)

This series of thermal cycles is carried out thanks to a programmable instrument, the thermal cycler, capable of changing the temperature very quickly and keep it constant for a given period. The result of a PCR is that, at the end of *n* cycles of amplification, the reaction mixture contains a theoretical maximum number of double-stranded DNA equal to *2n* (where "*n*" represents the number of amplification cycles).

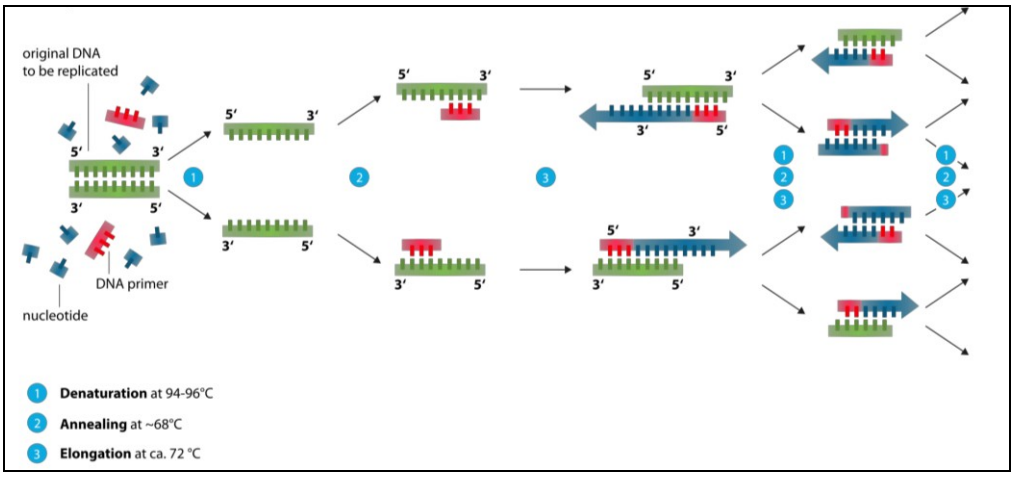

Figure 8: *PCR*: a schematic representation of PCR reaction is reported.

The reagents requested in a PCR are listed above:

1) DNA polymerase: different polymerases exist, and, according to the characteristics of the template and of the reaction, it is possible to select the most appropriate enzyme. One of the most used is represented by Taq polymerase. It derives from the *Thermophilus acquaticus* bacterium, it is stable at high temperatures and works with maximum efficiency between 72°-75°C. The thermal stability is a critical feature of this enzyme. Taq polymerase at 72°C has an enzymatic activity that allows the incorporation of 50-60 nucleotides per second, which corresponds, approximately, to 3 Kb per minute. The optimal enzymatic concentration spans from 0.5 to 2.5 U. A too high concentration may decrease the specificity of the reaction, while a too low concentration may not enable the conclusion of all cycles;

2) *Reaction buffer*: it is a Tris-HCl and KCl based buffer necessary to reproduce the optimal conditions for the activity of the polymerase thus increasing the throughput or the number of nucleotides that the enzyme can insert in succession before separating from the template strand;

3) Mg^{2+}: it is essential for the activity of Taq polymerase as its bond with the enzyme stabilizes it in a three-dimensional conformation that facilitates its activities. The Taq polymerase shows its highest activity around a concentration of free Mg^{2+} equal to 1.2-1.3 mM;

4) *dNTPs*: the solution of dNTPs contains the four bases of DNA: dATP, dGTP, dTTP, and dCTP. For a good efficiency of the PCR the four nucleotides must be present in equimolar concentration of around 50-200μM. A too high concentration may increase the incorrect rate of incorporation, while a too low concentration may reduce the reaction efficiency;

5) *Primers*: primers design is usually performed with specific software that facilitates their choice such as Primer3Plus. The aim of the primer design is to obtain a balance between amplification efficiency and specificity. Given a target DNA sequence, primer design software attempts to strike a balance between these two goals by using pre-selected default values for each of the primer design available. In particular, optimal primer pairs should be closely matched in Melting Temperature (Tm) and must not be able to form loops and primer dimers. Primer length (about 20-base pairs), sequence and GC contents are taken into account to select proper primers sequences.

PCR process depends on different factors, thus in genetic analysis a proper experimental design is necessary in order to obtain data.

SANGER DIRECT SEQUENCING

The "dideoxy method of DNA sequencing" was developed by Sanger and colleagues in 1977 (Sanger et al., 1977).

The enzyme used for this method is the AmpliTaq® DNA polymerase, a mutant form of Taq DNA polymerase with the following characteristics:

- it contains a point mutation in the active site that replaces the phenylalanine with a tyrosine at residue 667 (F667Y). This mutation results in a reduced capability to discriminate ddNTPs from dNTPs (Tabor and Richardson, 1995);
- it presents another point mutation in the amino terminal domain, replacing the glycine with an aspartate at residue 46 (G46D). This mutation removes almost all the 5′-3′ exonuclease activity, reducing the presence of possible artifacts;
- it has been formulated with a thermally stable inorganic pyrophosphatase that cleaves the inorganic pyrophosphate (PPi) byproduct of the extension reaction and prevents its accumulation in the sequencing reaction.

The sequencing procedure is based on the capability of the DNA polymerase to incorporate also 2′,3′-dideoxynucleotides (ddNTPs). These nucleotide analogues are associated with specific fluorophores, used to identify the different bases. Each dye emits light at different wavelengths when excited by an argon ion laser. Thus, all four colors and therefore all four bases can be detected and distinguished. When one of this unnatural ddNTP terminators is incorporated at the 3′ end of the growing chain, the elongation phase is interrupted due to the lack of the 3′-hydroxyl group.

To execute a direct sequence reaction, it is necessary to perform the following procedure better shown in *Figure 9*:

1. amplification of the desired fragment with PCR,
2. purification of PCR products,
3. second step of amplification with the proper mix of dNTPs and ddNTPs,
4. second step of purification of the obtained products,
5. separation of the fragments with capillary electrophoresis associated with laser detection.

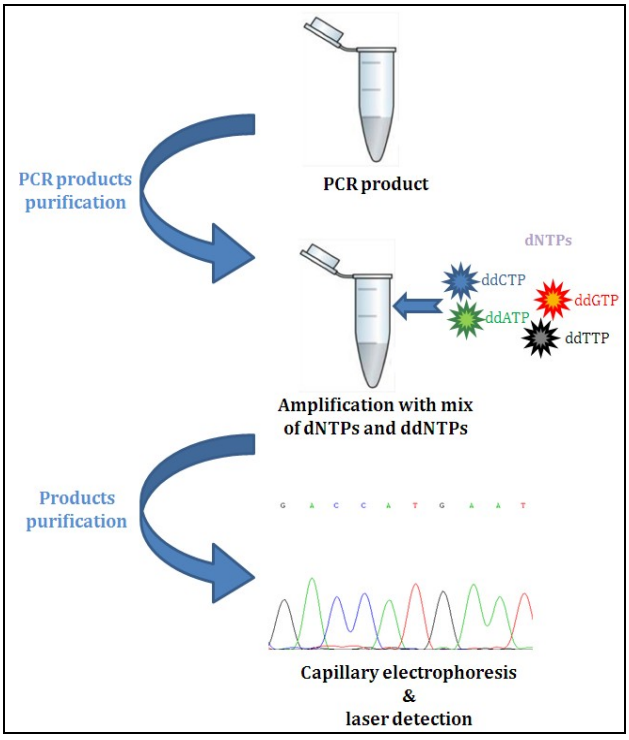

Figure 9: *Preparation of samples for Sanger sequencing*: a schematic representation of the procedure necessary to perform a direct sequencing is here reported.

After the amplification of the region of interest with PCR, the resulting products should be purified. PCR primers, dNTPs, DNA polymerase, and buffer components present in solution could indeed affect the performance of the sequencing reaction and lead to generation of noisy and non-specific data.

There are several methods for purifying PCR products:

- ✓ column purification
- ✓ enzyme purification (e.g. EXOsap)
- ✓ tips purification

In this thesis, we used the Rapid Diffinity Tip2 ® marketed by Sigma. These tips effectively remove dNTPs, primers, primer dimmers, and DNA polymerase while providing greater than 90% recovery of pure DNA fragments from 100bp to 10kb. The impurities are removed from the solution as they enter the pipette tip and, after mixing for just one minute, high quality DNA samples ready to use for downstream applications are obtained.

The purified DNA is then used for the direct sequencing. This reaction requires the use of a proper mix containing dNTPs and labeled ddNTPs.

For each sample, a 10µL mix was set-up containing 1.2µL buffer, 1µL BigDye (dNTPS, ddNTPs, DNA polymerase, 1µL sequence primer (0.33 µM), 4.8µL milliQ water, 2µL purified amplicon.

The amplification of the samples takes place following this thermal cycler program:

-96°C 1 min
-50°C 30 seconds } 30 cycles
-60°C 2 min

All the reactions start from the same nucleotide and end with a specific base, when the ddNTPs are added. Thus, in solution DNA chains with different lengths covering all the nucleotides' positions are obtained (Russell PJ, 2002).

After the amplification, unincorporated dye terminators have to be removed because they can interfere with base calling. Several protocols are currently used to purify these products. In this thesis, a simple and cheap precipitation method was applied based on the use of a resin-based protocol in 96-well plate format. This resin – generally hydrated superfine Sephadex-G50– retains salts, reactants, primers, and unincorporated dyes, while allowing the purified DNA to pass through this matrix during centrifugation. The purified samples are collected in a clean plate.

Once performed these reactions, DNA fragments are chemically denaturated with formammide and also thermally denatured with a rapid cycle of heating and freezing in ice. The samples are loaded in the Genetic Analyzer ABI Prism 3130 (Applied Biosystems, Foster City, CA), where a capillary electrophoresis associated with fluorescence detection happens. Since the four dyes emit fluorescence signals at different wavelengths, a laser reads the gel to determine the identity of each band. The results are then depicted in the form of a chromatogram, which is a diagram of colored peaks corresponding to the nucleotides. The obtained data are then analyzed using the free download software Chromas lite.

In this thesis, three SNPs (*SMAD3*-rs17228212, *Tudor-SND1*-rs3823994, *SMAD3*-rs744910) were genotyped by Sanger sequencing. In *Table 3*, the primers sequences and the PCR thermal profiles that have been setup for the analyses of these variants are reported.

	***SMAD3*-rs17228212**	***Tudor-SND1*-rs3823994**	***SMAD3*-rs744910**
Primer FRW (5'-3')	GCAGCACTTGGGCTAGTCAC	ATAGGGGGAGTCCGACACTT	TGGGAAGAAGTGAAGAAGAGG
Primer REV (5'-3')	GGTTCCAGCTCTTGGTTCTG	TCGAAAGCATCCCCTCTTAC	TCTGAAGCTTGGAATCAAAGAG
Thermal profiles	95°C for 5min 95°C for 30sec 62°C for 30sec } 35 cycles 72°C for 30sec 72°C for 10min	95°C for 5min 95°C for 30sec 59°C for 30sec } 35 cycles 72°C for 30sec 72°C for 10min	95°C for 5min 95°C for 30sec 60°C for 30sec } 35 cycles 72°C for 30sec 72°C for 10min

Table 3: *PCR conditions*: primers sequences and PCR thermal profiles setup for the analyses of *SMAD3*-rs17228212, *Tudor-SND1*-rs3823994, *SMAD3*-rs744910.

BEAD XPRESS READER COUPLED WITH VERACODE ® TECHNOLOGY AND GOLDENGATE ASSAY ®

Illumina BeadXpress Reader (Illumina, La Jolla, CA) is a dual-color laser scanning system allowing users to analyze several genetic markers in a multiplexing manner exploiting the VeraCodeTM microbeads digital technology. This technology allows several types of multiplex testing ranging from genotyping, gene expression, protein-based assays, and expression studies. This system can analyze from 1 to 384 biomarkers per sample at the same time. The VeraCode™ system is based on the use of VeraCode Beads, microcylinders (240µm in length by 28µm in diameter) inscribed with a unique digital holographic code to unambiguously designate and track the specific analyte or genotype of interest throughout the multiplex reaction. They are made by highly pure glass, stable at high temperatures and chemical agents, that represents an optimal surface for biomolecules attachment. In the GoldenGate Genotyping® Assay, each microbead is functionalized with a specific oligonucleotide which univocally identifies a single SNP. These microbeads are used in solution, allowing more rapid hybridization times.

To perform a BeadXpress analysis, a quite long preparation step is required as reported in *Figure 10*.

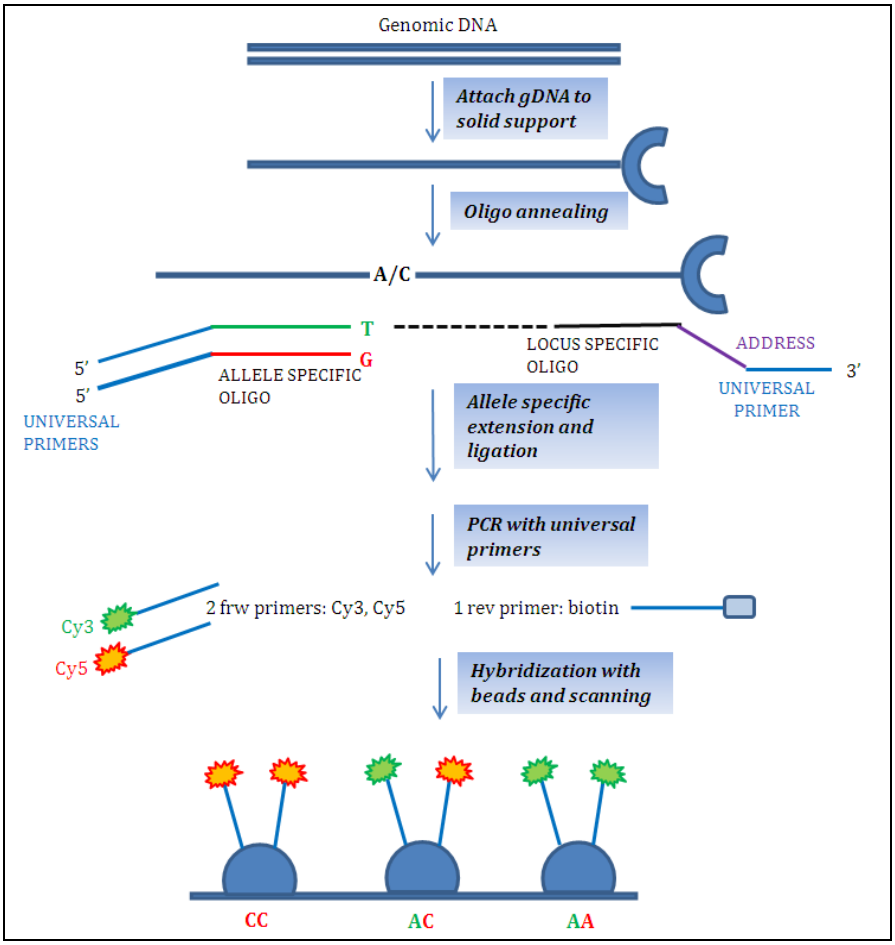

Figure 10: *BeadXpress analysis*: sample preparation steps for BeadXpress analysis are reported.

The first step in the GoldenGate Assay is the genomic DNA (gDNA) activation by biotinylation. This process enables the binding between gDNA and paramagnetic particles through biotin-streptavidin interaction. The activation process is highly robust and requires only 250ng of gDNA. Subsequently, gDNA is incubated with a mix containing different oligonucleotides, each conjugated with universal PCR primers. Specifically, three oligonucleotides are designed for each SNP locus: two are allele-specific oligonucleotides (ASOs), one is a locus-specific oligonucleotide (LSO). The ASOs differ only for the last nucleotide, which matches the polymorphic site on the gDNA. So, for each DNA strand, only one ASO hybridizes, depending on the SNP genotype. The LSO hybridizes several bases downstream from the SNP site. Moreover, the LSO contains a unique address sequence (or "Lumicode") that targets a particular oligonucleotide-coated VeraCode microbead type. Assay oligonucleotides (ASOs and LSOs), hybridization buffer, and paramagnetic particles are then combined with the activated DNA in the hybridization step. During the primer hybridization process, ASOs and LSOs hybridize to the gDNA sample bound to paramagnetic

particles. Because hybridization occurs prior to any amplification steps, no amplification bias is introduced into the assay. Following hybridization, several wash steps are performed, removing excess and mis-hybridized oligonucleotides. After that, the allele specific extension occurs, followed by ligation. Consequently, the ligation products join the information about the SNP genotype and the address sequence on the LSO. These products are isolated from gDNA after a step of denaturation. The single stranded ligation products serve as PCR templates. The multiplex PCR is performed using the universal PCR primers called P1, P2, and P3. Specifically, the P1 and P2 are used as forward primers. Thanks to their labeling with Cy3 and Cy5 dyes, the allelic discrimination is allowed. There is only one reverse primer, the P3, specific for the locus site. The P3 allows the amplification of the address sequence. This is fundamental for the binding with the bead specific for the analyzed SNP.

After downstream processing, the single-stranded, dye-labeled PCR products are hybridized to their complementary bead type through their unique address sequences. Hybridization of the GoldenGate Assay products onto the VeraCode beads separates the assay products for individual SNP genotype readout.

At the end, the BeadXpress® Reader is used for microbead code identification and fluorescent signal detection. Specifically, the plate is loaded in the BeadXpress Reader and beads from 8 wells at a time are drawn up and aspirated onto the 8-chambered transparent groove plate in which, thanks to a combination of fluid flow, gravity and capillary force, they align closely within the grooves. Once the beads are aligned, the optical system scans them. In particular, a dual-color laser detection system identifies both the unique holographic code of each VeraCode bead, specific for each SNP, and the fluorescence signal associated with each bead, discriminating the genotype. Assays developed with VeraCode microbeads typically include up to 30 replicates of each bead type. Each microbead is optically scanned up to a dozen times providing about 300 independent data point for each analyte ensuring reliable and accurate results.

The plate preparation process lasts about two days and the workflow is summarized in *Table 4*.

Process	Time	Day
DNA activation	1h 20 min	1
Oligonucleotides addition and DNA binding to paramagnetic particles	3h	1
Oligonucleotides-DNA binding	50 min	1
PCR amplification	2h 30 min	1
Amplicons isolation	1h 40 min	2
Amplicons hybridization with Veracode Bead	3h	2
Veracode Bead Plate washing	10 min	2
Veracode Bead Plate reading	1h 10 min-96 SNPs 3h 30 min-384 SNPs	2

Table 4: *BeadXpress workflow*: in this table we reported the workflow, the duration of the various steps, and the day when the various preparation steps are performed.

Data generated using the BeadXpress Reader are analyzed with Illumina's GenomeStudio™ data analysis software, which performs automated genotype clustering and calling. Specifically, the software associates the fluorescence data and the correspondent genotype. A clusterization algorithm assembles in three groups the fluorescence values related to each sample basing on the presence of only one (in case of homozygous genotypes) or two (heterozygous genotype) kind of signals. This process allows the determination of the genotypes for each investigated SNP. Each dot represents one sample (*Figure 11*).

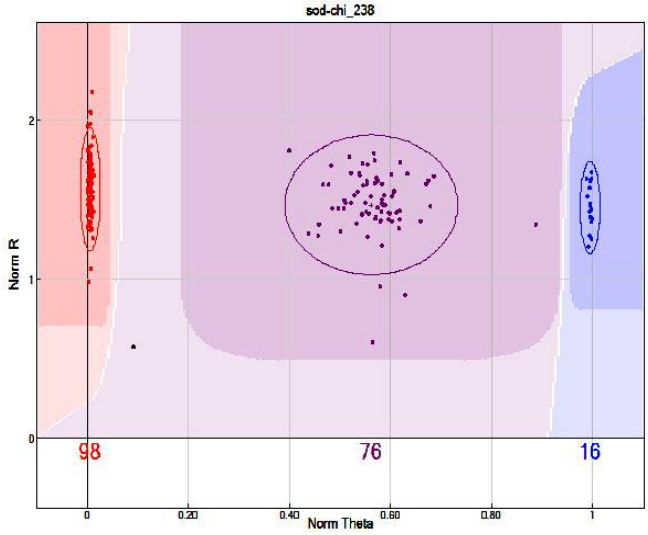

Figure 11: *Graphical representation of data obtained with BeadXpress analysis*: each dot represents an analyzed sample, the three colors are associated with the three genotypes: red AA, purple AB, blue BB.

The software automatically generates some scores, related both to sample and to analysis quality. According to these scores and to guidelines defined by the company, it is possible to manually optimize the results in order to obtain more robust data. In particular, for these projects we consider the following parameters:

1. *internal controls*: 48 sample-dependent, sample-independent, and contamination controls are all built into the GoldenGate assay. These controls provide a way to assess the overall performance of samples, reagents, and equipment. During preliminary sample quality evaluation, samples falling outside the expected performance parameters should be highlighted for additional analysis. Failure in these controls could be due to processing failures during sample processing or poor DNA quality;
2. *quality of analyzed samples*: before evaluating the quality of SNP clusters, it is important to highlight samples that have poor performance in the genotyping assay. The GenCall score is a quality score, ranging from 0–1, calculated for each genotype. GenCall scores generally decrease in value if the dot representing the sample is far from the center of its cluster;
3. *SNP cluster features*: to identify loci that should be manually edited or excluded, the software gives different scores, such as the cluster separation (that measures the separation among the three genotype clusters), the SNP call frequency (that considers the percentage of genotypes that have been determined in the analysis), and the intensity of the signals.

The identification of problematic samples and loci in a systematic manner ensures optimal final data quality from the GoldenGate genotyping assay.

4.4 STATISTICAL ANALYSES

During this thesis, two different projects were performed with different aims. Accordingly, two different statistical approaches have been performed. We exploited two different statistical programs:

- SAS 9.2 and Stata 11.2 (Stata-Corp, TX) for logistic and survival analysis
- "R" statistical package version 2.6.2 for Classification And Regression Tree (CART) analysis

4.4.1 IDENTIFICATION OF PREDICTIVE BIOMARKERS OF RESPONSE TO NEOADJUVANT TREATMENT

To define new potential predictive biomarkers of tumour response to neo-adjuvant treatment, TRG was used as the response parameter. Specifically, complete responders (TRG=1) were compared to partial or non responders (TRG=2-5). Patients were divided into two groups according to RT dose level of the neoadjuvant treatment (either 50.4Gy or 55Gy). A χ^2 test was applied to evaluate the differences in the distributions of demographic and clinical variables and treatment related factors between the two groups.

The association between genotypes and TRG were tested separately in the two groups of patients. The associations were adjusted according to gender, age, distance of the tumour from the anal margin, and time between the end of RT and surgery. Odds ratio (OR) and 95% confidence interval (95%CI) were computed. Dominant, recessive, and additive genetic models were considered for each genotype combining heterozygous and homozygous genotypes. The best fitting genetic model was selected according to the Wald X^2-test. Polymorphisms resulted significant in at least one group, showing a concordant genetic effect and with a compatible genetic model in the two groups were further investigated with a pooled analysis in the entire population. Results were validated by bootstrap analysis, fixing a total number of re-sampling of 1000. We considered as good predictive markers SNPs with p-value<0.05.

4.4.2 Definition of the interactions among the predictive genetic biomarkers and patients' clinico-pathological parameters

To investigate how genetic factors and clinico-pathological features interact to regulate neoadjuvant treatment response, a classification and regression tree (CART) analysis was performed. The process starts with the root node that contains all the responders (TRG=1; n=78) and the non-responders (TRG=2–5; n=202) subjects. At the end, a decision tree is obtained where the most important variables impacting the treatment response are highlighted.

4.4.3 Identification of prognostic genetic biomarkers

The prognostic value of the SNPs was testes evaluating different end-points. Specifically, we investigated the potential association between SNPs and:

1. the DFS, defined as the interval between surgery and relapse, death, or the last follow up;
2. the 2-year disease-free survival (2yDFS);
3. the 10-years OS, defined as the interval between surgery and the last follow-up.

Dominant, recessive, and additive genetic models were considered for each SNP combining heterozygous and homozygous genotypes. The best fitting genetic model was selected according to the Wald X^2-test. We considered as good prognostic biomarkers SNPs with p-value<0.05.

The effect of genotypes on DFS, 2yDFS, and OS was assessed through hazard ratios (HR) and the corresponding 95%CI. The survival analysis was computed with the Kaplan-Meier method, and multivariate Cox proportional hazard model was applied to test differences

between the two groups of patients defined according to their genotypes and the selected genetic model.

The identified significant associations between genetic variants and the 2yDFS and the OS were further validated with bootstrap analysis, fixing a total number of resampling of 1000. We considered as good prognostic biomarkers SNPs with p-value<0.05.

5. RESULTS

5.1 PATIENTS' ENROLMENT, TREATMENT, AND CLINICAL DATA COLLECTION

Patients' clinical and pathological data were collected from the medical records. According to the elegibility criteria, 280 patients were selected and analyzed for our purposes. Patients' clinico-pathological parameters are reported in *Table 5.*

	N	(%)
PERSONAL DATA AT DIAGNOSIS		
GENDER		
Male	191	68.21
Female	89	31.79
AGE (years)		
average, range	61 (20-82)	
TUMOUR DISTANCE FROM ANAL MARGIN (cm)		
<8	189	67.50
≥8	91	32.50
NEOADJUVANT TREATMENT		
RADIOTHERAPY DOSAGE (Gy)		
50.4	202	72.14
55.0	78	27.86
FLUOROPYRIMIDINES ADMINISTRATION		
bolo	15	5.36
continuous infusion	98	35.00
os	155	55.36
not available	12	4.29
PLATINUM ADMINISTRATION		
yes	102	36.43
no	175	62.50
not available	3	1.07
TUMOUR REGRESSION GRADE (TRG)		
1	78	27.86
2	58	20.71
3	90	32.14
4	43	15.36
5	11	3.93
ADJUVANT TREATMENT		
yes	133	47.50
no	132	47.14
not available	15	5.36
RELAPSE		
yes	62	22.14
no	218	77.86

Table 5: *Patients' and treatment parameters*: in this table patients' baseline characteristics and details about treatment are reported. Gy: Gray.

5.2 BIOINFORMATIC ANALYSIS : SNPS SELECTION

We defined two different bioinformatic approaches to select two panels of SNPs involved in the pathways of interest. These approaches were meant in order to join literature data with the great potentialities offered by bioinformatic tools freely available on line.

5.2.1 SNPS SELECTION : MIRNA PROJECT

According to the strategy better described in 4.2.1 paragraph, we identified 144 SNPs localized in 51 genes involved in miRNAs activity and maturation analyzable with BeadXpress technology. In particular, 117 SNPs were localized in miR-machinery factors and 27 in miRNAs possibly involved in the translational control of POLR2A, Drosha, DGCR8, and Dicer (*Figure 12)*.

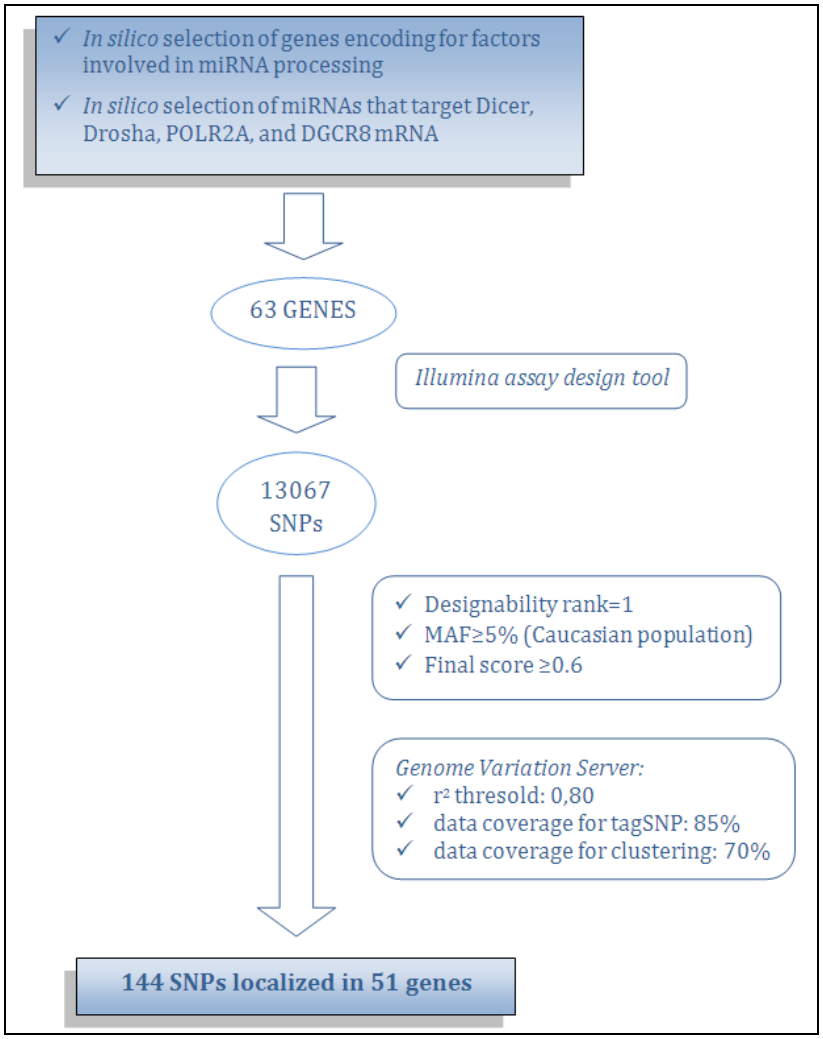

Figure 12: *Selection of genes and SNPs*: a strategy joining bioinformatic and literature analyses was performed to select a panel of 144 SNPs involved in miRNA activity and maturation. MAF: minor allele frequency.

5.2.2 SNPs SELECTION : IMMUNOGENETICS PROJECT

A second SNPs selection was performed to identify SNPs potentially involved in immune response activity. Specifically, according to the strategy better described in 4.2.2 paragraph, we identified 192 SNPs localized in 34 genes involved in immune system and cancer and analyzable with BeadXpress (*Figure 13*).

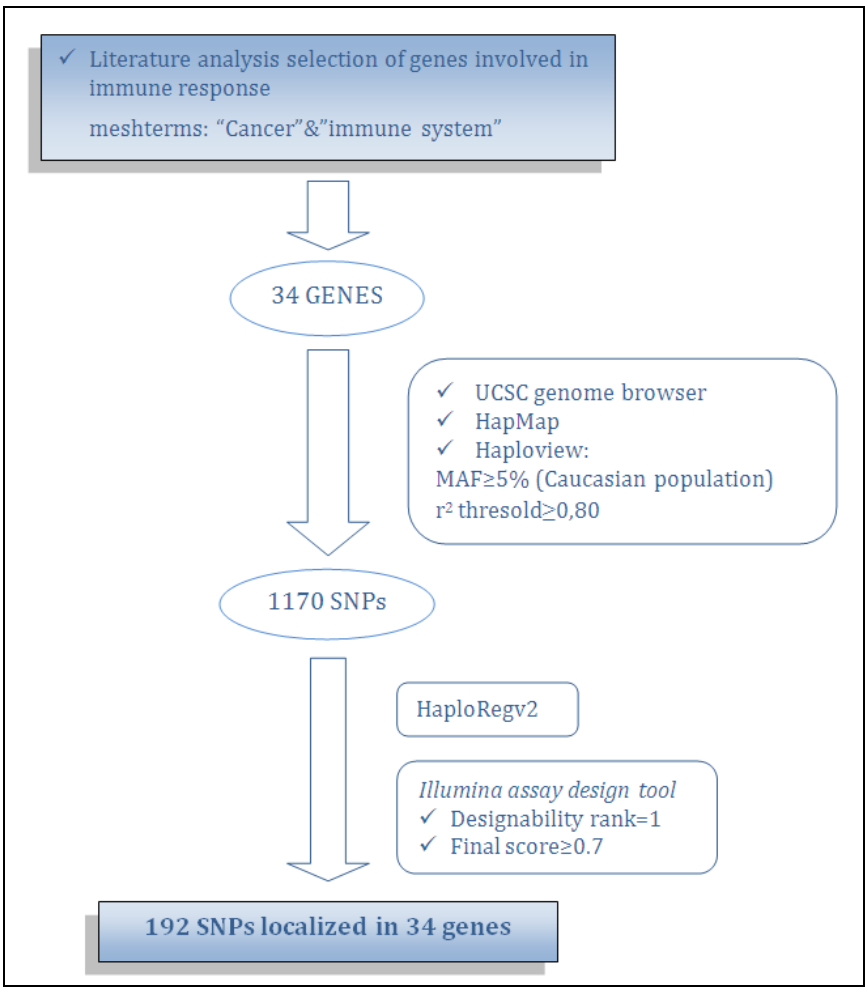

Figure 13: *Selection of genes and SNPs*: a strategy joining bioinformatic and literature analyses was performed to select a panel of 192 SNPs correlated with immune system and cancer. MAF: minor allele frequency.

5.3 MOLECULAR ANALYSES

The genetic analyses of germinal DNA samples of the eligible LARC patients were performed with the panel of 144 SNPs potentially involved in miRNA activity and maturation. Genotyping analyses were successful in 114 assays out of 144. Fifteen DNA samples were excluded from genotyping due to their poor quality. The average genotype call rate of these analyses was 98.4% (range: 95.8-100.0%). Three SNPs were randomly selected for

BeadXpress analytical validation by Sanger sequencing. In particular, 93 samples were sequenced for *SMAD3*-rs17228212, 62 for *SMAD3*-rs744910, and 99 for *Tudor-SND1*-rs3823994. All these SNPs showed a high concordance rate (100.00% *SMAD3*-rs17228212, 100.00% *Tudor-SND1*-rs3823994, 98.38% *SMAD3*-rs744910). The successful analytical validation ensured the high reliability of BeadXpress results.

The panel of 192 SNPs involved in immune system was also used for LARC patients' analyses. Genotyping analyses were successful in 147 assays, 45 need to be redesigned. Forty-five samples were excluded from genotyping due to their low quality, so at the end we obtained genotypes of 235 patients. The average genotype call rate was 93.36% (range: 70.64-99.57%). Positive controls were included in the analyses and the 100% concordance rate ensured the reliability of these analyses.

The percentage of successful assays in the miRNA panel was 79.2%, whereas the percentage of successful assays in the immunogenetics panel was 76.6%. The quite low success rate of these molecular analyses can be ascribed on the one hand to the not optimal quality of the employed DNA samples and on the other hand to the quite strict criteria applied for the elaboration of the results. These criteria led to exclude many assays but ensured the high reliability of the results that was assessed with both the analyses of positive controls and with the direct sequencing of some samples.

5.4 STATISTICAL ANALYSES

5.4.1 IDENTIFICATION OF PREDICTIVE BIOMARKERS OF RESPONSE TO NEOADJUVANT TREATMENT

One primary aim of this thesis was the definition of new potential predictive biomarkers of tumour response to neo-adjuvant treatment. For this purpose, we analyzed the panel of SNPs associated with miRNA activity and maturation.

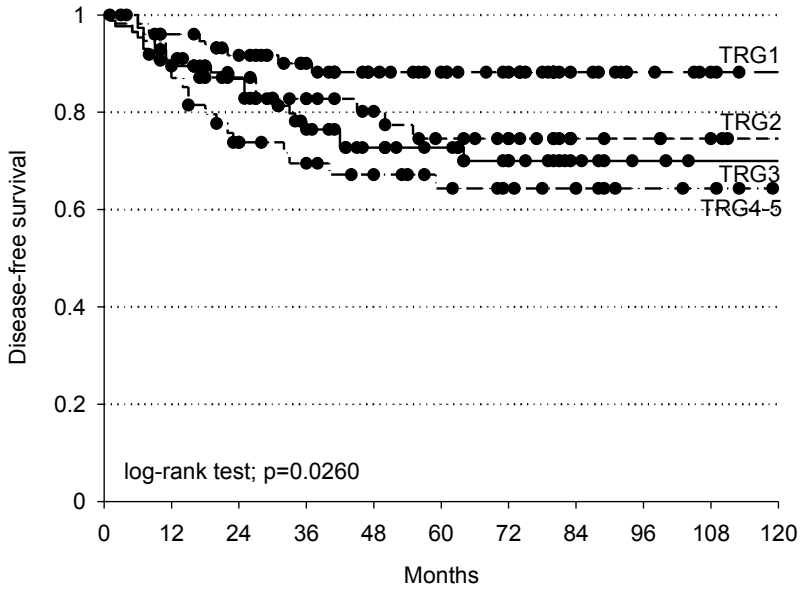

Figure 14*: Kaplan-Meier estimates of DFS according to TRG:* black dots represent censored patients. Log rank test was applied to compare DFS between complete responders (TRG=1) and non-complete responders (TRG=2-5). TRG: tumour regression grade, DFS: disease free survival.

The TRG was used as response parameter. Specifically, patients obtaining a pathological complete response (pCR), defined as TRG=1, were compared to non responders (TRG=2-5). This classification was defined because DFS was significantly different between these two groups of patients (log-rank test p=0.0260) (*Figure 14*).

We defined a study flow chart aimed at identifying predictive biomarkers not affected by the different RT dosage administered in the neoadjuvant setting. Study design is better described in *Figure 15*.

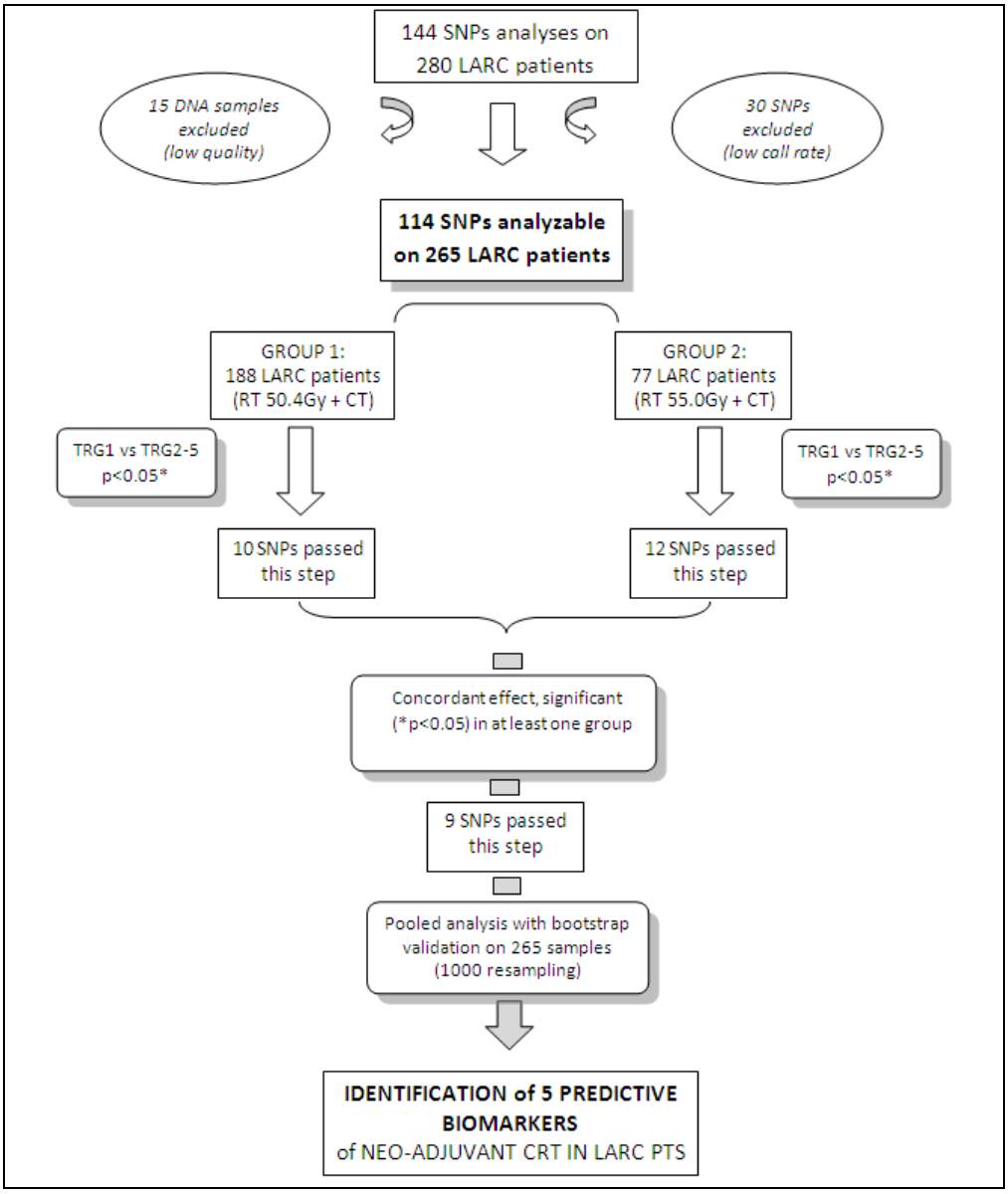

Figure 15: *Study flow chart*: the image represents the flow chart of the study performed to identify predictive biomarkers of response to neoadjuvant treatment. LARC, Locally Advanced Rectal Cancer; SNPs, single nucleotide polymorphisms; TRG, Tumour Regression Grade; RT, Radiotherapy; CT, Chemotherapy; Gy, Gray; CRT, Chemoradiotherapy; Pts, patients; CART, Classification And Regression Tree; DFS, Disease Free Survival. *Fisher exact test.

Patients were divided into two groups according to RT dose level of the neoadjuvant treatment (either 50.4Gy or 55.0Gy). Patients were well balanced for sex, age, and distance of the tumour from the anal margin (*Table 6*).

	RT 50.4Gy* (188 patients)	RT 55.0Gy (77 patients)	X^2
PERSONAL DATA AT DIAGNOSIS			
GENDER			
Male	55 (29.3%)	28 (36.4%)	
Female	133 (70.7%)	49 (63.6%)	p=0.2573
AGE (years)			
<55	47 (25.0%)	19 (24.7%)	
55-59	37 (19.7%)	13 (16.9%)	
60-64	35 (18.6%)	20 (26.0%)	
65-69	36 (19.2%)	12 (15.6%)	
>70	33 (17.6%)	13 (16.8%)	p=0.7264
TUMOUR DISTANCE FROM ANAL MARGIN (cm)			
<8	124 (66.0%)	54 (70.1%)	
≥8	64 (34.0%)	23 (29.9%)	p=0.5114
NEOADJUVANT TREATMENT			
TUMOUR REGRESSION GRADE (TRG)			
1	53 (28.2%)	20 (26.0%)	
2	36 (24.7%)	20 (26.0%)	
3	58 (15.5%)	29 (37.7%)	
4-5	41 (21.3%)	8 (10.4%)	p=0.1153

Table 6: *Distribution of patients, according to treatment (radiation therapy dose) and clinical features*.
*93 out of 188 patients (49.5%) received oxaliplatin in addition to fluoropyrimidines.

The association between genotypes and pCR was tested separately in the two subgroups of patients treated at different RT dose, with multivariate analysis (*Table 7*). Ten SNPs resulted significant (p≤0.05) in the 50.4Gy group, twelve in the 55.0Gy group. Nine SNPs (*DROSHA*-rs10719, *TRBP*-rs6088619, *SMAD3*-rs17228212, *SMAD3*-rs744910, *SMAD3*-rs745103, *SMAD5*-rs1057898, *SMAD5*-rs6871224, *TNRC6A*-rs6497759, *miR-371a*-rs28461391) resulted significant in at least one group, showed a concordant genetic effect, and compatible genetic models in the two subgroups. Considering the concordant, RT-dose independent effect of the previously identified nine SNPs, we performed a pooled analysis of the combined datasets to increase the statistical power, and then we internally validated the results by bootstrap analysis.

Five SNPs were still significantly associated to pCR (*Table 8*). *DROSHA*-rs10719 and *SMAD3*-rs17228212 were associated with an higher risk to be non-complete responders (TRG=2-5) according to an additive model (OR=1.87, 95%CI=1.10-3.17, p=0.0274; and OR=2.01, 95%CI=1.22-3.31, p=0.0049 respectively). The same effect was observed for *SMAD3*-rs744910 and *SMAD3*-rs745103, according to a recessive model (OR=0.45, 95%CI=0.24-0.85, p=0.0153; and OR=0.48, 95%CI=0.25-0.94, p=0.0471, respectively). *TRBP*-rs6088619 showed an opposite effect, according to an additive model (OR=0.39, 95%CI=0.19-0.79, p=0.0125).

			RT 50.4 Gy (n=188)			RT 55.0	
Gene	**SNP**	**Base change**	**GM**	**OR (95% CI)[a]**	**p-value**	**M**	**OR (95**
CNOT4	rs11772832	A>G	R	1.65 (0.65-4.23)	0.2937	R	**0.11 (0.0**
CNOT6	rs6877400	A>G	R	0.28 (0.02-4.83)	0.3785	D	**0.16 (0.0**
DDX20	rs197412	A>G	**A**	**1.83 (1.05-3.21)**	**0.0339**	D	0.73 (0.2
DGCR8	rs417309	G>A	D	1.73 (0.54-5.59)	0.3553	D	**0.20 (0.0**
DICER1	rs1057035	A>G	**D**	**2.25 (1.07-4.72)**	**0.0327**	R	0.62 (0.1
DROSHA	rs10719	C>T	**A**	**2.39 (1.24-4.61)**	**0.0094**	R	2.74 (0.2
TRBP	rs6088619	A>G	**A**	**0.34 (0.15-0.75)**	**0.0073**	D	**0.21 (0.0**
SMAD2	rs1792671	G>A	**D**	**0.16 (0.04-0.63)**	**0.0087**	R	0.29 (0.0
SMAD3	rs17228212	T>C	**A**	**1.83 (1.02-3.30)**	**0.0446**	A	**3.61 (1.1**
SMAD3	rs2289791	C>A	**A**	**0.58 (0.35-0.97)**	**0.0364**	R	4.60 (0.3
SMAD3	rs744910	A>G	R	0.50 (0.23-1.07)	0.0739	R	**0.16 (0.0**
SMAD3	rs745103	A>G	R	0.59 (0.27-1.27)	0.1819	A	**0.23 (0.0**
SMAD3	rs8025774	G>A	**D**	**0.47 (0.24-0.92)**	**0.0279**	R	4.33 (0.3
SMAD3	rs8028147	G>A	D	1.60 (0.83-3.11)	0.1609	A	**0.31 (0.1**
SMAD5	rs1057898	A>G	D	0.71 (0.36-1.39)	0.3157	D	**0.12 (0.0**
SMAD5	rs6871224	A>G	D	0.72 (0.37-1.43)	0.3485	D	**0.10 (0.0**
TNRC6A	rs6497759	T>A	D	1.33 (0.65-2.73)	0.4368	D	**6.63 (1.0**
TNRC6B	rs139911	T>C	**A**	**1.71 (1.04-2.80)**	**0.0353**	D	0.12 (0.0
miR196A2	rs11614913	C>T	**R**	**0.29 (0.11-0.78)**	**0.0138**	D	0.44 (0.1
miR371A	rs28461391	C>T	D	0.92 (0.41-2.04)	0.8288	D/A	**0.20 (0.0**

Table 7: *First selection of SNPs*: only the associations between SNPs and TRG (TRG2-5 vs TRG1) according to RT dose sig
of patients are reported. Statistically significant association are in bold. SNP, single nucleotide polymorphism; RT: radiot
A: additive; D: dominant.

Genes	SNP	Base change	Genotype frequency						M	OR (95%
			TRG 1			TRG 2-5				
			AA	Aa	aa	AA	Aa	aa		
DROSHA	**rs10719**	**C>T**	0.616	0.370	0.014	0.492	0.450	0.058	A	1.87 (1.10
TRBP	**rs6088619**	**A>G**	0.747	0.239	0.014	0.863	0.137	0.000	A	0.39 (0.19
SMAD3	**rs17228212**	**T>C**	0.708	0.264	0.028	0.524	0.377	0.100	A	2.01 (1.22
SMAD3	**rs744910**	**A>G**	0.219	0.438	0.343	0.289	0.529	0.185	R	0.45 (0.24
SMAD3	**rs745103**	**A>G**	0.219	0.493	0.288	0.277	0.559	0.165	R	0.48 (0.25
SMAD5	rs1057898	A>G	0.366	0.507	0.127	0.500	0.385	0.115	D	0.61 (0.34
SMAD5	rs6871224	A>G	0.366	0.507	0.127	0.497	0.392	0.111	D	0.61 (0.35
TNRC6A	rs6497759	T>A	0.754	0.217	0.029	0.649	0.309	0.042	D	1.70 (0.89
miR371A	rs28461391	C>T	0.718	0.268	0.014	0.783	0.201	0.016	D	0.66 (0.34

[a]Adjusted for gender, age, distance from anal margin, time between radiation therapy and surgery, platinum treatment, a

Table 8: *Genetic predictive biomarkers of response to neoadjuvant treatment*: this table reports associations between the pathological complete response (TRG2-5 vs TRG1) performed in the pooled population (n=265) and with Bootstrap cor reported in bold (p<0.05). SNPs, single nucleotide polymorphisms; TRG, Tumour Regression Grade; RT, radiotherapy; Gy Interval; M, genetic model; R, recessive; A, additive; D, dominant.

5.4.2 DEFINITION OF THE INTERACTIONS AMONG THE PREDICTIVE GENETIC BIOMARKERS AND PATIENTS ' CLINICO -PATHOLOGICAL PARAMETERS

To better define the role of SNPs and clinico-pathological features in the response to neo-adjuvant treatment, the five significant SNPs and some clinico-pathological features (gender, age, RT dose, kind of neo-adjuvant treatment, time between the end of CRT and surgery, and distance of the tumour from the anal margin) were tested in the CART analysis. As reported in *Figure 16*, each terminal node identifies a specific combination of genetic and clinico-pathological features that is associated with a different probability to completely respond to neo-adjuvant treatment.

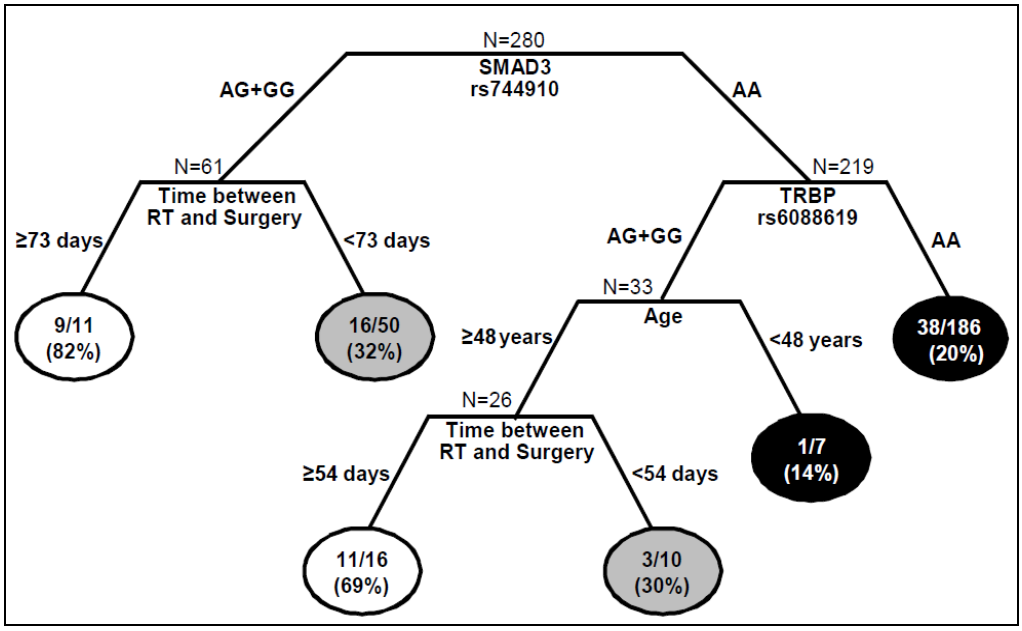

Figure 16: *CART analysis*: CART analyzing SNPs and clinico-demographic characteristics predictive of pathological complete response is here reported (pCR)(TRG=1). Terminal nodes report the number and fraction of patients who achieved pCR. White circles represent terminal nodes with high probability of achieving pCR, grey circles represent terminal nodes with intermediate probability of achieving pCR, black circles represent terminal nodes with low probability of achieving pCR. RT, Radiotherapy.

The first node is determined by *SMAD3*-rs744910. According to the recessive model, patients were split in those carrying at least one variant allele (AG+GG) and those homozygous for the wild type allele (AA). The first group was further stratified according to the time between the end of CRT and surgery. The longest interval (≥73 days) was correlated to the highest probability of pCR (82%). If the interval is shorter than 73 days, the probability to be good responders decreased (32%). Among patients with *SMAD3*-rs744910 AA genotype, *TRBP*-rs6088619 AA genotype was associated with a very low probability to be good responders (20%), while for patients carrying at least one G allele, age acted as discriminator. Young patients (<48 years) were associated with a low

probability of pCR (14%). In patients with ≥48years of age, the time between the end of CRT and surgery discriminated complete and non-complete responders.

This analysis highlighted two subgroups of patients associated with high probability to achieve a pCR (82% and 69%), two with an intermediate probability (32% and 30%), and two with low pCR probability (14% and 20%), according to their genetic characteristics and clinico-pathological features.

5.4.3 IDENTIFICATION OF PROGNOSTIC GENETIC BIOMARKERS

Another primary aim of this thesis was the definition of new potential prognostic biomarkers of LARC patients. For this purpose, we analyzed both panels of SNPs and we searched for association between SNPs and prognosis defined in terms of DFS, 2yDFS, and OS.

GENETIC BIOMARKERS ASSOCIATED WITH DFS

The SNPs showing a robust predictive value identified in the miRNA panel (specifically *DROSHA*-rs10719, *TRBP*- rs6088619, *SMAD3*-rs17228212, *SMAD3*-rs744910, and *SMAD3*-rs745103) were also investigated for their possible association with DFS. The results were adjusted according to gender, age, distance of the tumour from the anal margin, time between the end of CRT and surgery, RT dosage, type of surgery, and adjuvant CT. The best fitting genetic models identified in the analyses regarding treatment response were applied also to compute the HR and the corresponding 95%CI.

There was no significant associations between any of the investigated SNPs and DFS. Only a trend ($p<0.10$) between *TRBP*-rs6088619 and DFS can be highlighted: patients with at least one G allele had a lower probability to have a bad DFS (HR=0.24, 95% CI=0.07-0.79, p=0.0706) (*Figure 17*).

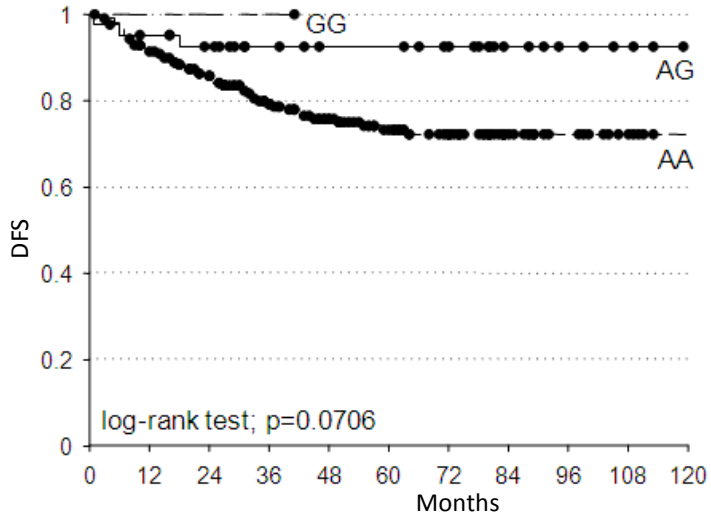

Figure 17: *Kaplan-Meier estimates of DFS according to TRBP-rs6088619*. Black dots represent censored patients.

However, an interesting observation come from the analysis of patients with complete pathological response (TRG=1): in this subgroup *SMAD3*-rs745103 was significantly associated with DFS (p=0.011). In particular, patients bearing the GG genotype have a significant better DFS then those carrying at least one A allele (*Figure 18*).

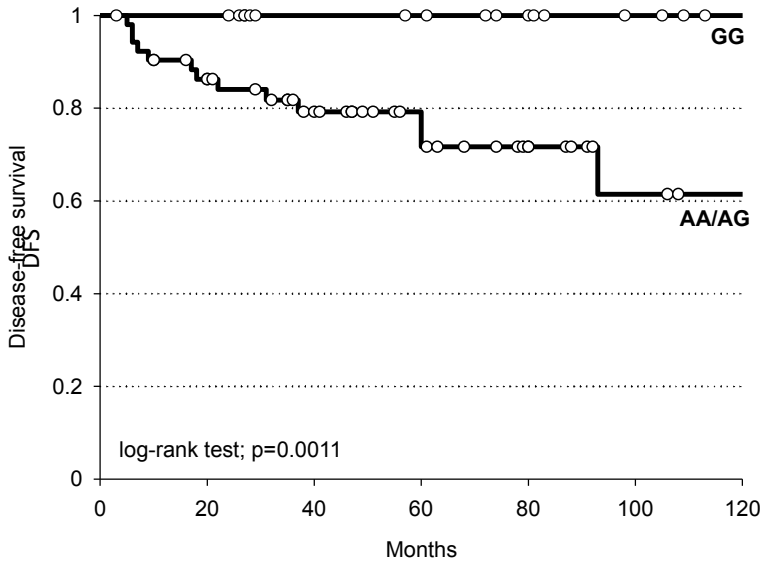

Figure 18: *Kaplan-Meier estimates of DFS according to SMAD3-rs745103 in patients with TRG=1.* Dots represent censored patients.

GENETIC BIOMARKERS ASSOCIATED WITH 2YDFS AND OS

We analyzed the prognostic value of SNPs involved in immune response in terms of both 2yDFS and OS. A multivariate analysis was performed adjusting data for age, gender, RT dosage, kind of surgery, adjuvant treatment. Results were therefore tested with bootstrap analysis.

We identified four genetic markers significantly associated with 2yDFS by multivariate logistic regression which were still significant after bootstrap validation, according to a dominant genetic model. Three of them were risk factors, *IL17F*-rs641701, *IL17F*-rs9463772, and *TGFβ receptor 2* (*TGFβ R2*)- rs9867701 (OR=5.84, 95%CI=1.52-22.45, p=0.010; OR=3.56, 95%CI=1.22-10.35, p=0.020; and OR=3.00, 95%CI=1.09-8.30, p=0.034, respectively). Moreover, also a protective marker arose, *STAT3*-rs8069645 (OR=0.36, 95%CI=0.13-0.99, p=0.048). We further analyzed the prognostic value of these biomarkers evaluating their association with the 10-year OS. According to the dominant model, three of them resulted significantly associated to 10-year OS by multivariate COX regression. Specifically, *IL17F*-rs641701 and *IL17F*-rs9463772 were risk factors for bad prognosis (HR=3.23, 95%CI=1.50-6.95, p=0.003 -*Figure 19*-, and HR=2.89, 95%CI=1.49-5.61, p=0.002 -*Figure 20*-, respectively). Patients homozygous for the G allele for either *IL17F*-rs641701 or *IL17F*-rs9463772 had a better prognosis then the other ones. Moreover, *STAT3*-rs8069645 was significantly associated with 10-years OS. Patients with at least one G allele have a higher probability to experience a good prognosis (HR=0.50, 95%CI=0.25-0.98, p=0.044 -*Figure 21*-).

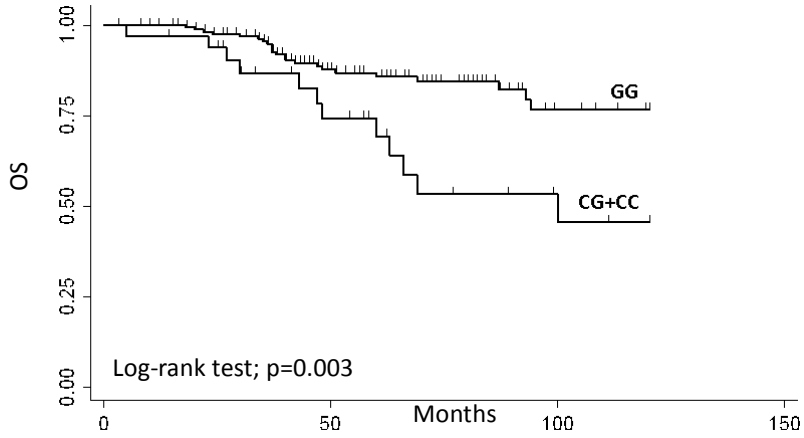

Figure 19: *Kaplan-Meier estimates of OS according to IL17F-rs641701.* Dots represent censored patients.

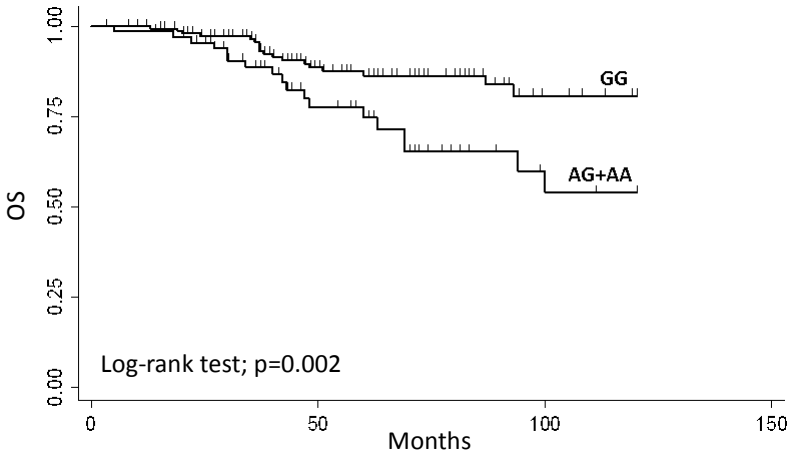

Figure 20: *Kaplan-Meier estimates of Os according to IL17F-rs9463772*. Dots represent censored patients.

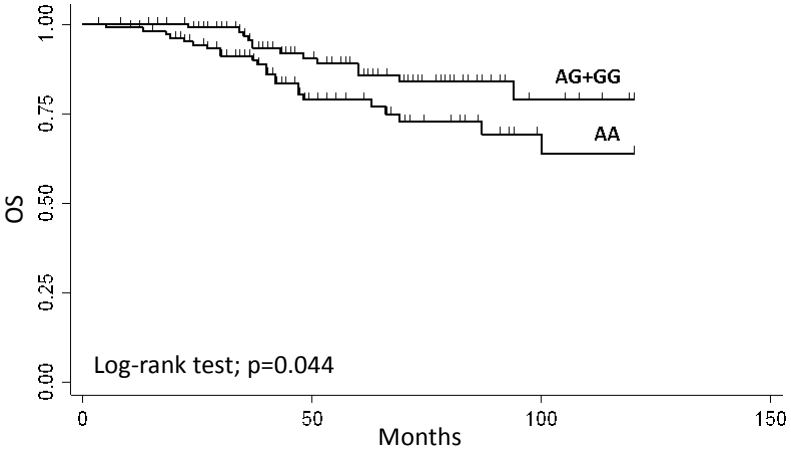

Figure 21: *Kaplan-Meier estimates of OS according to STAT3-rs8069645*. Dots represent censored patients.

We were able to confirm the prognostic value of these biomarkers in a validation group of 63 LARC patients who underwent radical surgery and adjuvant treatment based on fluoropyrimidines sometimes associated with platinum derivatives.

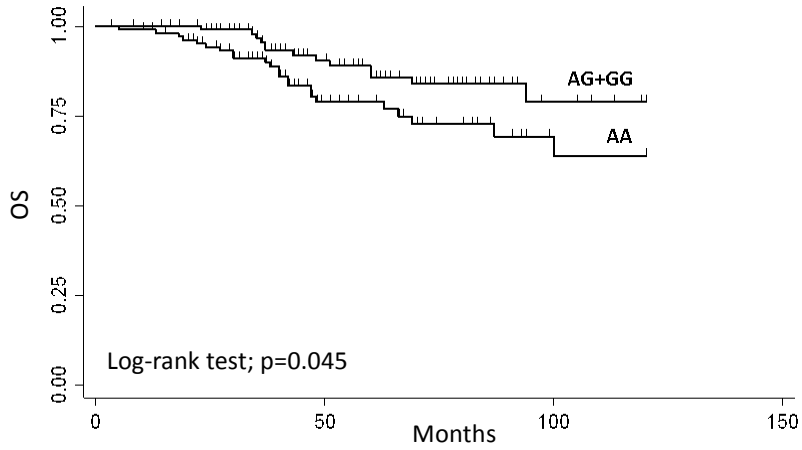

Figure 22: *Kaplan-Meier estimates of OS according to IL17F-rs9463772.* Dots represent censored patients.

The SNP *IL17F*-rs9463772 was significantly associated with OS (p=0.045). As reported in *Figure 22*, patients bearing at least one G allele have a higher probability to have a good prognosis compared to patients homozygous for the A allele.

We can thus confirm that this is really a strong prognostic biomarker.

6. DISCUSSION

PGx represents a discipline with big potentialities in the field of personalized medicine that for sure in the future will play a pivotal role in cancer patients' management. Despite the introduction of some PGx biomarkers in the international guidelines of some drugs, patients and clinicians are still waiting to reap the benefits potentially associated with genetics. As understanding treatment response can be of crucial importance, a more comprehensive approach, open to new and still unexplored scenarios is demanded.

The need to broad the fields of interest of PGx studies is closely examined in this work. In particular, during this thesis, we explored the potential clinical role of genetic variants in genes involved in miRNA maturation and immunity. The importance of these pathways is on the rise and is sustained by an increasing number of literature evidences. We investigated the potentialities of the PGx of miRNAs and of immunogenetics in LARC patients. In particular, we addressed two clinical problems related to this malignancy: the response to neoadjuvant treatment and patients' prognosis. Finding an answer to these questions could be translated into a significant improvement for patients, both for treatment personalization and for an optimized follow-up.

The primary aim of this thesis was to identify potential predictive biomarkers to neoadjuvant treatment through the analysis of 114 SNPs selected for their potential involvement in miRNA maturation and activity on 265 LARC patients.

It is estimated that more than 30% of the human genes are post-transcriptionally regulated by miRNAs, highlighting their capability to deeply influence the majority of the pathways (MacFarlane and Murphy, 2010). Since their discovery, more than 21,000 publications have addressed the association between miRNAs and cancer (PubMed-analysis done on the 15 th March 2016). The miRNAs impact the expression level of several known drug-related genes, thus playing a direct role in treatment response. This is fuelling the research on this field, in particular on the big potentialities of the analysis of miRNA-related SNPs (Sclafani et al., 2015; Ganzinelli et al., 2015).

Specifically, the genetic analyses applied in this thesis were meant to identify new potential predictive biomarkers of response to neoadjuvant CRT performed during LARC treatment. This represents a compelling need as no general consensus has been reached up to date on the proper management of these patients and only 30% reach a complete pathological response to this treatment (Aschele et al., 2005; Probst et al., 2015). Thus, the identification of patients who have a low probability to respond to therapy can be of great interest as they could avoid the neoadjuvant treatment, sparing useless toxicities related to CRT and

optimizing time for the surgical intervention. Moreover, treatment optimization will ultimately play a pivotal role in the prognosis (Maas et al., 2011).

For this reason, we considered a complete pathological response (TRG=1) as the clinical end-point of the study, since, consistently with our data (*Figure15*), this represents the most reliable prognostic factor nowadays. Moreover, we designed a statistical strategy aimed to identify SNPs that can predict the response to neoadjuvant CRT independently of the applied RT dosage. This was conceived in order to select markers that can be potentially applied to different schedules of treatment.

We identified five new potential predictive biomarkers of response to neoadjuvant treatment. Three of them are located in *SMAD3* (rs17228212, rs744910, and rs745103), one in *Drosha* (rs10719), and one in *TRBP* (rs6088619).

SMAD3 is a transcriptional modulator that can directly activate Drosha, a nuclear RNAse III that catalyzes the first cleavage of pri-miRNAs. Specifically, after being activated by cytokines like TGFβ, SMAD3 directly binds a consensus sequence located in specific pri-miRNAs, as miR-21, promoting their processing by Drosha complex (Davis et al., 2008; Davis et al., 2010). Intriguingly, miR-21 seems to affect both prognosis and response to treatment of cancer patients. Its overexpression was associated with node positivity, metastasis, and poor survival in CRC patients (Kulda et al., 2010; Slaby et al., 2007). Moreover, in two independent cohorts of advanced CRC patients, the correlation between miR-21 expression levels and response to 5-FU-based adjuvant treatment was demonstrated (Schetter et al., 2008). SMAD2/3 complex directly induces also the transcription of some miRNAs, like miR-192 and miR-451(Chung et al., 2010; Gal et al., 2008), that are involved in CRT response (Skinner et al., 2014; Bandres et al., 2009).

Consistently with our data, the phosphorylated SMAD3 overexpression in pre-CRT cancer tissues was recently correlated with poor response to fluoropyrimidines-based neoadjuvant CRT of 86 LARC patients (Huang et al., 2015). Interestingly, that study referred to TRG as response parameter as we did, so the obtained results are comparable. Both studies strongly highlight the key role played by SMAD3 in neoadjuvant response, and both genetic analyses and protein expression can be of great relevance in neoadjuvant CRT optimization in LARC patients.

A further potential predictive biomarker is *TRBP*-rs6088619. TRBP is a dsRNA-binding protein involved in RISC assembly and in Dicer activity. The stability of TRBP-Dicer complex affects the selection of pre-miRNAs that will be processed.

These findings highlight the pivotal role of miRNA maturation in the response to neoadjuvant treatment in LARC. However, even if miRNA-related genes represented the

main focus of our study, we must consider that SMAD3, Drosha, and TRBP are involved in other cellular pathways.

SMAD3 is a transcriptional modulator in TGFβ pathway and regulates many related-cancer genes like *SERPIN1* and *FOXA2* (Zhang Y et al., 2011). Intriguingly, TGFβ induces via SMAD2/3 complex the transcription of *NADPH oxidase 1* and *NADPH oxidase 4*, up-regulating the production of reactive oxygen species, thus potentially enhancing RT efficacy (Hubackova et al., 2015). Drosha, on the other hand, is involved in DNA repair system regulating the maturation of both miRNAs and other ncRNAs. Specifically, ATM and BRCA1, master players of cell response to DNA damage, once activated, can promote the Drosha-mediated processing of specific pri-miRNAs such as let-7a1, miR-16a, miR-145, and miR-34a (Zhang X et al., 2011; Kawai and Amano, 2012).

Drosha is also involved in the synthesis of another class of ncRNAs called *DNA damage response RNAs* that directly mediate DNA repair (d'Adda di Fagagna, 2014). The involvement of Drosha in the DNA repair system could be of major importance when considering our clinical model and the CRT response.

Taken together, these data are really intriguing because, even if the aforementioned pathways are not directly related to miRNA maturation, they could be of major relevance to understand our results.

We believe that one of our major findings is the identification of the key role of SMAD3 in treatment response, highlighted by the fact that three of the five significant SNPs are located in three different haploblocks of this gene. We can thus assume that these variants support, in an independent way, the importance of SMAD3.

These SNPs are intronic and, at the best of our knowledge, their biological function has not been elucidated yet. Literature data are available only for *SMAD3*-rs745103: its robust prognostic role was assessed in colon- but not in rectal cancer (Slattery and Lundgreen, 2014). Interestingly, also in our analysis the prognostic role of this SNP was assessed. In particular, it was significantly associated to DFS in patients with complete pathological response (TRG=1). This result, if confirmed in future analyses, could have a pivotal clinical impact. Usually, patients with complete pathological response do not undergo adjuvant treatment because they are at low risk to relapse. As you can see from the *Figure 18*, this SNP can stratify patients according to their risk to relapse, information that can be translated into clinics prescribing for these patients more frequent visits during the follow-up or, eventually, with the administration of CT also after surgery.

We further investigated the potential effect of these SNPs exploiting bioinformatics tools. Specifically, we analyzed them with the "*SNP function prediction tool*" available on SNPinfo

web server, without obtaining significant results. Thus, we studied also SNPs located in the haploblocks of rs17228212, rs744910, and rs745103, as defined by GenomeVariationServer referring to Caucasian population. Five SNPs (rs731874, rs10152307, rs10152987, rs11630297, and rs11634560) were identified to be in linkage with rs17228212, three with rs744910 (rs10152544, rs11634793, and rs12708492), whereas no SNPs were in linkage with rs745103. Neither in these cases functional predictions were obtained.

Another intronic SNP resulted significant from our analysis is *TRBP*-rs6088619. There are no evidences in literature about its function that is not predictable with the aforementioned bioinformatic approach. Moreover, GenomeVariationServer did not suggest any SNP in linkage with it.

Regarding these intronic SNPs without any known and predictable function, we can mention projects like Encode that are shedding light to the importance of non coding regions in regulating gene expression (Maher, 2012). Consequently, we cannot exclude that these intronic SNPs could exert a still unknown regulatory role in gene transcription or translation that maybe will be clarified in the future.

The predictive SNP highlighted for *Drosha* (rs10719) is located in the 3'UTR and, according to microSNiper, TargetScanHuman, and miRanda, it affects the binding of Drosha mRNA with miR-181b. At the best of our knowledge, few literature data are available about this SNP. The only study investigating its potential predictive role failed: *Drosha*-rs10719 was indeed not successfully associated with response in metastatic CRC patients treated with 5-FU and irinotecan (Boni et al., 2011). Even if its function is not still experimentally determined, we can hypothesize that, through miRNAs regulation or with the involvement of other factors that bind 3'UTR, this SNP could alter Drosha mRNA stability or translation, affecting tumour response to CRT.

In an attempt to study how *SMAD3*, *Drosha*, and *TRBP* genetic variants interact with clinico-pathological characteristics of the patients we performed a CART analysis. The first node of the CART is *SMAD3*-rs744910, representing the most discriminating factor. Among patients that carried AA genotype, *TRBP*-rs6088619 seems to better further stratify patients according to their chance to get a TRG1. Age was the only non-genetic characteristic to be significant in the CART analysis. In a genetically defined subgroup of patients (AA according to *SMAD3*-rs744910 and AG/GG according to *TRBP*-rs6088619), young people (<48 years) had a higher risk to have a bad response to therapy. This could be explained by the more aggressive nature of cancers with an early onset (Chang et al., 2012), even if no correlations between age and tumour response have never been addressed before.

A noteworthy differential effect of an "actionable" clinical variant, the interval between the end of neoadjuvant CRT and surgery, was highlighted in two different subgroups. Based on our data, it would be advisable to wait more than 73 days for patients with AG/GG genotype for rs744910 *SMAD3* and more than 54 days for those bearing AA genotype for rs744910 *SMAD3* and AG/GG genotype rs6088619 *TRBP* and older than 48 years. This information can be of particular interest since, actually, there is no consensus on the proper timing for surgery after CRT and no clear effect of prolonging this interval on patients response and prognosis was obtained (Foster et al., 2013; You et al., 2015). Considering the contrasting suggestions coming from literature, it can be advisable to personalize the interval between the end of neoadjuvant treatment and surgery, maybe stratifying patients according to their genetic profile as we propose with our study. However, these results must be carefully verified and examined in prospective studies.

To conclude, with the analysis of the SNPs of the miRNA panel we identified five potential predictive biomarkers of response to neoadjuvant treatment and, thanks to the CART analysis, we also highlighted the key role of the time interval between the end of CRT and surgery in determining the pathological response. However, it goes without saying that the fully understanding of the factors involved in the response to neoadjuvant treatment in LARC patients could have extraordinary clinical significance and the validation of the obtained results could be translated in the introduction in the everyday clinical practice of new genetic biomarkers.

During this thesis we explored another field of study of PGx that is represented by immunogenetics. Immune system plays an active role in different stages of cancer onset, from its burning to the formation of metastasis, exerting a direct activity both in the microenvironment and in the global response to cancer due to the plethora of cytokines and immune mediators usually found in the blood circulation. Cancer progression is indeed a complex process involving host–tumour interactions through multiple molecular and cellular factors of the tumour microenvironment. The pivotal connection between immune system and cancer establishment is robustly explained by many scientific evidences (Grivennikov et al., 2010) and has fostered the conceivement of a so called "*immunoscore*" that could be a potential tool for clinicians in the future (Galon et al., 2012). Moreover, another investigated field of study is represented by the analysis of the potential clinical role of SNPs located in genes involved in immunity, that is immunogenetics. Specifically, there are already literature data addressing the potential prognostic role of some SNPs in immune-related factors, that can be explained by the potential impact they exert on the global inflammatory state of the patients (Schoenfeld et al., 2013; Markkula et al., 2014).

These literature data, even if still not conclusive, corroborated the second part of this thesis. In particular, we aimed at analyzing the potential prognostic role of a panel of SNPs in immune-related genes on LARC patients. Patients' prognosis, and, in particular, the recurrence after radical surgery, represents another clinical problem related with this malignancy. Most of the recurrences (80%) occurs within the first 4 years after surgery, so the identification of prognostic biomarkers could promote a better follow-up with, hopefully, an improved disease control (Valentini et al., 2015). The heterogeneous evolution of this disease claims the identification of intermediate endpoints of OS. One of them is represented by the already analyzed pathological complete response, which was the focus of the first project. Another possible surrogate endpoint of OS is the two-year disease free survival (2yDFS). The group of Valentini published an interesting paper that compares these two endpoints for OS in LARC patients, analyzing a pooled dataset including data obtained by five large European randomized trials. They identified several risk groups of patients joining these two indicators. Intriguingly, they highlighted the importance of the immunological response in tumour growth. The 2yDFS seems to have a stronger impact on OS compared to pCR.

This study prompted us to select the 2yDFS as another endpoint of our analyses. Specifically, we analyzed 147 SNPs in immune-related genes on 235 LARC patients and we searched for their potential association with the 2yDFS. The resulting significant SNPs were thereafter tested for their association with the OS.

Our analyses led us to identify 4 SNPs significantly associated with the 2yDFS that were also internally validated with bootstrap analysis. Two of them are located in *interleukin 17F* (*IL17F*) (rs641701 and rs9463772) and the other ones in *signal transducer and activator of transcription 3* (*STAT3*) (rs8069645 and rs9867701). We further analyzed their prognostic value testing their association with the 10 years OS. Interestingly, 3 SNPs remained significant: *IL17F*-rs641701, *IL17F*-rs9463772, and *STAT3*-rs8069645. We were able to test these associations in a validation group of 63 LARC patients who underwent radical surgery and adjuvant treatment based on fluoropyrimidines sometimes associated with platinum derivatives. It is important to note that, *IL17F*-rs9463772 was still significantly associated with OS (p=0.045), thus this approach unveiled the pivotal prognostic role of this genetic biomarker.

The obtained results are consistent with the literature data available nowadays about the role exerted by IL17F and STAT3 in cancer setting.

STAT3 is an oncogene that regulates many pathways. The presence of cytokines like those with proinflammatory activity released by inflammatory cells infiltrating the tumour bulk,

like IL6, promotes the activity of Tyr kinases as Janus kinase (JAK), that, in turns, activate STAT3. Specifically, after being phosphorylated, it forms a dimer and translocates in the nucleus, where it regulates the transcription of several genes. Specifically, its activation induces cell proliferation, resistance to apoptosis, and angiogenesis. Persistent activation of STAT3 characterizes different human malignancies, as CRC (Qin et al., 2015), since it is demonstrated that it promotes tumorigenesis, highlighting its key role also in the primary stage of cancer development (Nguyen et al., 2014). Moreover, the high tumour expression of this protein is significantly associated with poor prognosis in CRC patients (Morikawa et al., 2011). In light of these findings, the activity of some chemotherapeutic molecules on this pathway was analyzed, and it was found that camptothecin and oxaliplatin promote apoptosis *in vitro* blocking the activation of STAT3 mediated by IL6 (Cross-Knorr et al., 2013). Interestingly, a recently published study demonstrated that the resistance to 5-FU in CRC can be explained by the overexpression of *cyclin D1*, that is the main regulator of the progression from G1 to S phase of the cell cycle. The promoter of this cyclin hosts a STAT3 binding site. Intriguingly, in CRC tumoral tissues the overexpression of *cyclin D1* is correlated with the overexpression of *STAT3*, that in turns promotes resistance to cisplatin (Huerta et al., 2003) and to 5-FU (Qin et al., 2015). Regarding this, inhibition of STAT3 leads to sensitization of cells to 5-FU based CRT, highlighting another interesting link between this factor and our clinical model (Spitzner et al., 2014). Moreover, literature data underline the key role of this factor in the epithelial-mesenchymal transition also in CRC, reinforcing our results on the association of STAT3 in disease local control (Rokavec et al., 2014) (Lee et al., 2014).

All these findings shed light the potentiality to conceive STAT3 as a new druggable target. Two analyses had already suggested the use of a natural inhibitor of STAT3, scoparone, that is extracted from *Artemisia capillaris*. Both studies demonstrated an antiproliferation activity of this molecule, even if in different settings (Park et al., 2015). If validated, these results could pave the way to a new kind of cancer treatment.

In this study, the most important result is represented by the SNPs located on *IL17F*. This factor belongs to a big family of interleukins which also includes IL17A, IL17B, IL17C, IL17D, IL17E. All these proteins are well conserved among mammals.

IL17F is mainly secreted by lymphocytes T-helper 17 (Th17), even if other cell types produce this cytokine, like neutrophils, monocytes, and natural killer cells (Cua and Tato, 2010), broadening the spectra of mechanisms in which IL17 is involved. Interestingly, IL17F mRNA was also seen in colonic epithelial cells (Ishigame et al., 2009), suggesting that this factor can also modulate immune response by the involvement of non lymophoid cells.

Once released, IL17F forms a homo- or a heterodimer with IL17A. The complex can bind to both IL17 receptor A (IL17RA) and IL17 receptor C (IL17RC). IL17RA is commonly expressed in immune cells, whereas IL17RC in non immune ones. It is worth of note that IL17F presents an higher affinity to Il17RC, suggesting again the importance also of non immune cells in the response orchestrated by IL17F. The binding to the receptor promotes the downstream activation of different signaling pathways involving Raf-1, mitogen-activated protein kinase (MEK) 1 and 2, and extracellular signal-regulated kinase (ERK) 1 and 2.

More in detail, IL17F can promote the expression of proinflammatory cytokines such as tumour necrosis factor (TNF), IL1, and the already cited IL6, exacerbating the inflammatory process. This cytokine is noteworthy in the oncological setting due to the role it plays in various phenomena. For instance, IL17F was found to inhibit the angiogenesis of endothelial cells and, of note, *IL17F* gene is located on chromosome 6, in a genomic region that includes also *VEGF*, the master regulator of this process. Additionally, IL17F can induce endothelial cells to produce IL2, and TGF-β, and it can promote the production of matrix metalloproteinases by fibrobasts, endothelial cell, and epithelial cells. These enzymes are in turn involved in tumour invasion and metastatization.

Due to the plethora of functions that regulates, it is not surprising that IL17F has been studied not only in autoimmune disease and inflammatory conditions but also in the oncological setting. Many groups have already associated this cytokine with CRC (Zeng et al., 2015; Omrane et al., 2015). In CRC human tumour specimens, *IL17F* is down-regulated, impacting cancer development. *In vivo* studies demonstrated that its overexpression plays an inhibitory role in cancer development, that can be explained through the angiogenesis inhibition promoted by IL17F (Tong et al., 2012).

Taken together, the data obtained thus far about the role of IL17F in cancer are quite puzzling and more efforts are requested to clarify the contributions of this cytokine in immune response and cancer.

To better interpret our findings, we searched for a link between STAT3 and IL17F and, interestingly, they are both involved in Th17 maturation and activity. IL17F acts in Th17 maturation and activity, whereas STAT3 is involved in the differentiation of Th17 and it is also an effector of pathways regulated by different interleukins including IL17. Many groups have analyzed the potential clinical role of Th17 infiltrates in cancer bulk. We can refer to an interesting study published by the group of Tosolini where a group of patients with CRC were enrolled and the tumoral inflammatory state based on the presence of Th17 and Th1 was assessed. Patients with high expression of genes associated with Th17 were at higher

risk of bad DFS compared to patients with high expression of genes related with Th1, who had a lower risk to relapse (Tosolini et al., 2011). This work was very recently commented by the group of Amicarella, that, through an extensive analysis on 1148 CRC samples, suggested, on one hand, the importance of a more comprehensive approach based on the analysis of all the cell populations producing IL17 and, on the other hand, the necessity to move with caution when hypothesizing target treatments against IL17/Th17 due to the dual role they exert in cancer (Amicarella et al., 2015). Moreover, it seems that Th17 function can also be affected by the type of cancer and the therapeutic approach (Zou and Restifo, 2010), paving the way to further interesting scenarios that need to be explored.

To better analyze the obtained genetic data, we performed a bioinformatic and literature analysis on the SNPs which resulted significantly associated with both the 2yDFS and the OS (*STAT3*-rs8069645, *IL17F*-rs641701, and *IL17F*-rs9463772). The performed approach is analogous to the one described for the study of the predictive biomarkers related with miRNA maturation and activity.

The *STAT3*-rs8069645 is intronic and, till now, its biological role has not been determined. Nonetheless, it was included in a panel of SNPs analyzed in the already cited paper published by Slattery, and a significant association with colon cancer mortality arose. Specifically, subjects carrying at least one G allele present a lower risk to experience a bad prognosis (Slattery and Lundgreen, 2014). This result is really interesting because, even if the analyzed clinical model differs from ours, the protective role of the G allele is highlighted and the obtained data are comparable in both studies.

To better investigate the potential biological impact of this SNP on the encoded gene, we performed a bioinformatic analysis with *SNPinfo web server* of the SNP of interest and of those in linkage with it in the Caucasian population ($r^2 \geq 0.8$). No predictions about *STAT3*-rs8069645 were obtained. Seven SNPs (rs11079043, rs3785898, rs4103200, rs4796793, rs6503696, rs6503697, rs9912773) are in linkage with *STAT3*-rs8069645 and one of them, *STAT3*-rs4796793, seems to affect the binding of different transcription factors such as HNF4α and PPARγ. This prediction, even if fascinating because it could explain the obtained result conferring a role in gene expression control of *STAT3*-rs8069645, needs to be verified with at least an *in vitro* analysis.

We analyzed the SNPs in *IL17F* gene with the same approach obtaining interesting results. The *IL17F*-rs641701 is a downstream variant of *IL17F*. At the best of our knowledge, no literature data about this SNP is available and, moreover, no functional predictions were obtained. According to SNPinfo, seven SNPs are in linkage with *IL17F*-rs641701: rs11965530, rs11968115, rs13209590, rs13218661, rs2245502, rs2397084, and

rs6913472. Only for *IL17F*-rs2397084 a prediction was available. It is indeed located in the coding region of the gene and it causes an amino acid substitution in position 126 (Glu126Gly). This SNP was very recently associated with oral squamous cell carcinoma risk (Li et al., 2015). The scarce available data cannot bring further hypothesis about *IL17F*-rs641701 role, even if its association with a non synonymous SNP is really interesting and could shed light also to the obtained result.

The most interesting result obtained in our study is the prognostic role of *IL17F*-rs9463772. In our analysis it represents the only genetic biomarker which prognostic role is confirmed in another group of LARC patients. This is a really important result considering that the lack of an external validation is one of the main limits of the introduction of PGx biomarkers in the everyday clinical practice. At the best of our knowledge, no literature data about the role of this SNP in cancer are available till now. We investigated the role of this SNP with our bioinformatic approach. It is located upstream of the gene codifying for IL17F, and, according to SNPinfo, it can affect the binding with transcription factors such as PPARγ. Moreover, it is in linkage with other SNPs (rs12201582, rs1266824, rs2294834, rs3857604, and rs7741835), and one of them, *IL17F*-rs2294834, is predicted to affect the binding with transcription factors too. Taken together, these preliminary data shed light to the potential gene expression regulation exerted by *IL17F*-rs9463772, paving the way to an interesting biological hypothesis that needs to be verified.

To sum up, with the analysis of the SNPs of the immunogenetic panel, three prognostic SNPs were identified. One of them is situated in *STAT3*, whereas the other ones are located on *IL17F* gene. The last SNPs are not in linkage, so they underline the potential key role of IL17F as prognostic biomarker. This hypothesis is reinforced by the validation of the prognostic role of *IL17F*-rs9463772 in another group of LARC patients. The obtained results are of course preliminary but, if validated, could represent a useful tool for clinicians to optimize LARC patients' management.

To go one step further, an overview about all the obtained results is necessary. This is fundamental considering the strict correlation between treatment response and prognosis, that represents the link between the miRNA project and the immunogenetics project. Specifically, we studied the biological mechanisms involving the genes highlighted in these two projects that are *SMAD3*, *TRBP*, and *Dicer* in the miRNA analysis and *IL17F* and *STAT3* in immunogenetics. Remarkably, two pathways seem to be in common: the TGFβ- and the MAPK ERK pathway.

To go in details, TFGβ can activate the SMAD2/3 complex (Song et al., 2009) and regulate Th17 maturation, that involves both STAT3 and IL17 (Rèbè C et al, 2013). Moreover,

according to recent literature data, STAT3 can also directly interact with SMAD3 antagonizing TGFβ signalling (Wang G et al, 2015).

We searched for possible associations between TFGβ and the analyzed clinical model. Consistently with our hypothesis, using an *in vitro* model able to test the sensitivity to a 5-FU-based CRT on 12 CRC cell lines, TGFβ was seen to impact the sensitivity to CRT (Salendo et al., 2013). Moreover, TGFβ exerts pro-angiogenic activities and mediates vessel stabilization, mechanisms known to be of primary importance in RT response (van Meeteren et al., 2011). Accordingly, serum TGFβ levels has shown to predict tumour response to neo-adjuvant 5-FU-based CRT in esophageal squamous cell carcinoma (Cheng et al., 2014).

The second pathway in common between the two projects is orchestrated by ERK, a kinase that regulates cell cycle, apoptosis, and DNA damage response, mechanisms that are crucial in CRT response. When ERK is activated, it catalyzes the phosphorylation of TRBP increasing the stability of the complex formed by TRBP and Dicer. This induces an increase in oncomiRNAs levels and a reduced expression of the oncosuppressor Let7 (Paroo et al., 2009). The activation of MAPK ERK can be catalyzed also by IL17F: after the binding with its receptor, this cytokine induces the activation of signalling cascades including also ERK. These results are fascinating, considering the direct involvement of this kinase in treatment response. Specifically, a better pathological response to 5-FU-based CRT was significantly associated with higher phosphorylated ERK levels in tumour specimens of rectal cancer patients (Davies et al., 2011).

Taken together, these literature data foster the involvement of TFGβ and ERK in LARC. The presence of these pathways in both projects is hypothesis generating because they can form the light motif between them. Hence, it will be of interest to investigate the role of these factors in our clinical model, studying genetic variants on *TFGβ* and *ERK* or analyzing their expression levels.

Nonetheless, we must acknowledge that this study presents some limits. Regarding the identification of predictive biomarkers among miRNA-related SNPs, we did not have the possibility to test them in an independent validation set of patients. To overcome this problem, we performed two independent analyses in two subgroups of patients treated with a different RT dosage and we selected only those presenting a concordant effect in both groups. From a statistical point of view, finding the same association with a concordant effect in two different groups of patients lowers the chance of false positive discoveries. Moreover, in an attempt to perform an internal validation we have applied a bootstrap re-sampling strategy considering only the results remaining significant after correction.

Another limit of the study is the lack of information on the biological function of the predictive SNPs. However, the location of these genetic variants in non-coding regions and the emerging data on the relevance of non-coding region in the regulation of gene expression, let us to presume a possible effect of these variants at the mRNA/protein level. In our study, the predictive effect of the genetic markers on pathological tumour response translated in an only marginal, non-significant difference in term of DFS. This is consistent with our hypothesis of the involvement of the markers on the local tumour response to RT, although we could hypothesize that other clinical and molecular variables could cooperate with TRG in determining patients' prognosis. In the present study, at pathologic examination of the surgical specimen, pCR was observed in 27.5% of the cases. This value is in the higher range of reported complete pathological response rates, and could be explained by the use in our patients of factors shown to be associated with complete pathological response after neoadjuvant treatment such as 50.4Gy or higher, continuous infusion of 5-FU, and two drugs regimens (Sanghera P et al, 2008). It is likely that the tumour response to CRT is a complex phenotype with a biological basis that probably depends on a plethora of tumour and host factors. Several previous studies tried to address the issue of pathological response to pre-operative treatment in rectal cancer. Both molecular (Hur et al., 2014; Brettingham-Moore et al., 2011) and radiological (Beets-Tan and Beets, 2014; Martens et al., 2015) measures have been evaluated, but none of them provided a reliable marker to be introduced by itself as selection criteria for patients' treatment (Ryan et al., 2015). Probably only a multi-parameters predictor will definitely address the issue of pCR in rectal cancer. In this context host genetic characteristics must be considered as one of the key players.

The second aim of this thesis was to identify potential prognostic biomarkers analyzing a set of SNPs in immune-related genes. With this project we provided data that corroborate the hypothesis that SNPs in genes involved in immune system might have an impact on patients' prognosis. We also had the opportunity to test the prognostic value of the identified biomarkers in a validation set, and through this analysis the prognostic value of *IL17F*-rs9463772 was further reinforced. Even if the result obtained for *IL17F*-rs9463772 is undoubtedly fascinating, further analyses in a larger group of validation and that also consider clinical variables are needed to corroborate our evidence. Unfortunately, we were not able to validate the prognostic value of the other identified SNPs in the external group of patients, even if their value seemed to be quite strong because reinforced with the bootstrap re-sampling analysis. This can be due to the quite small group of validation, that can highlight only strong prognostic biomarkers. Further analyses performed in larger group of

patients are mandatory to definitively clarify the prognostic value of these SNPs. Another limit is represented by the scarce information available till now about the biological function of the identified SNPs. We tried to give some suggestions through an elementary bioinformatic approach, that did not provide us many insights. However, we believe that it is really amazing that also in the immunogenetic panel non coding variants resulted significant and this, as already mentioned before, could stress once again the importance of these quite neglected regions. Future functional analyses of these genetic variants will for sure give the key to better understand their potential clinical role.

Further studies in large group of LARC patients and in prospective analyses are demanded in order to validate the clinical value of the genetic variants highlighted in our results. Moreover, other strategies could be of interest, such as the analysis of the expression level of the *SMAD3*, *TRBP*, *Dicer*, *IL17F* and *STAT3*, at local or at systemic level, according to the specific features of these proteins. Another possible strategy is to investigate the genetic variants arisen in our study also in other clinical models due to the broad spectra of mechanisms regulated by miRNAs and by immune system.

To conclude, in this thesis we explored the potentialities of two innovative PGx fields represented by SNPs potentially affecting miRNA activity and those involved in immune system activity. We applied these quite unexplored pathways to find new genetic biomarkers that can be translated in the optimization of LARC patients' treatment and management. Specifically, with the miRNA project we identified 5 new potential predictive biomarkers of response to neoadjuvant treatment (rs17228212, rs744910, and rs745103 on *SMAD3*, rs10719 on *Drosha*, and rs6088619 on *TRBP*) and, interestingly, we also highlighted the importance of an adjustable clinical parameter, that is the interval between the end of neoadjuvant treatment and surgery. The immunogenetic panel was applied to identify new potential prognostic biomarkers. Three SNPs were significantly associated with 2yDFS and OS, *IL17F*-rs641701, *IL17F*-rs9463772, and *STAT3*-rs8069645. Interestingly, the prognostic value of *IL17F*-rs9463772 was further reinforced with the analysis performed in an external group of validation.

All these results, if confirmed in future analyses, could be of great medical interest and may hopefully lead the optimization of LARC patients' treatment and management. Moreover, these results underline the great potentialities offered by pathways not directly involved in drug response that could represent the key for a deeper comprehension of different clinical problems.

7. CONCLUSIONS

In this thesis we explored the big potentialities offered by pathways not strictly related with drugs mechanism of action in PGx studies. We were indeed able to identify quite robust genetic biomarkers with high clinical potentialities in LARC patients' management.

Specifically, with the analysis of the miRNA panel, we identified five SNPs significantly associated with the pathological tumour response (TRG): three are located in *SMAD3* (rs17228212, rs7449140, and rs745103), one in *DROSHA* (rs10719), and one in *TRBP* (rs6088619). Moreover, the CART analysis highlighted the pivotal influence exerted by the time interval between the end of neoadjuvant therapy and surgery on treatment response in specific subgroups of patients. This finding is really intriguing due to the possibility to modulate this clinical parameter.

The immunogenetic panel underlined three new prognostic biomarkers significantly associated with 2yDFS and OS: *IL17F*-rs641701, *IL17F*-rs9463772, and *STAT3*-rs8069645. Interestingly, the prognostic value of *IL17F*-rs9463772 was validated in another group of 63 LARC patients.

In this thesis the use of bioinformatic tools enriched our analyses and shed light to the enormous potentialities of these tools in PGx studies. It is advisable to broad their application in order to optimize genetic analyses and to exploit the huge number of data that are already freely available online.

To conclude, with this thesis we obtained results with a big potential clinical impact. If confirmed in further studies, they will help clinicians to personalize neoadjuvant treatment of LARC patients and optimize their follow-up. We strongly believe that a deeper study of pathways not strictly related with drug activity and the exploitation of new tools like bioinformatics could be the key to a deeper comprehension of treatment response that ultimately affects patients' prognosis.

8. BIBLIOGRAPHY

Almeida, Maria Ines, Rui Manuel Reis, and George Adrian Calin. "Decoy Activity through microRNAs: The Therapeutic Implications." *Expert Opinion on Biological Therapy* 12, no. 9 (September 2012): 1153–59. doi:10.1517/14712598.2012.693470.

Amicarella, F., M. G. Muraro, C. Hirt, E. Cremonesi, E. Padovan, V. Mele, V. Governa, et al. "Dual Role of Tumour-Infiltrating T Helper 17 Cells in Human Colorectal Cancer." *Gut*, December 30, 2015. doi:10.1136/gutjnl-2015-310016.

Aravin, Alexei, Dimos Gaidatzis, Sébastien Pfeffer, Mariana Lagos-Quintana, Pablo Landgraf, Nicola Iovino, Patricia Morris, et al. "A Novel Class of Small RNAs Bind to MILI Protein in Mouse Testes." *Nature* 442, no. 7099 (July 13, 2006): 203–7. doi:10.1038/nature04916.

Aschele, C., M. L. Friso, S. Pucciarelli, S. Lonardi, L. Sartor, G. Fabris, E. D. L. Urso, et al. "A Phase I-II Study of Weekly Oxaliplatin, 5-Fluorouracil Continuous Infusion and Preoperative Radiotherapy in Locally Advanced Rectal Cancer." *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 16, no. 7 (July 2005): 1140–46. doi:10.1093/annonc/mdi212.

Bai, Lu, Hong Yu, Hetong Wang, Hongxin Su, Junhua Zhao, and Yuxia Zhao. "Genetic Single-Nucleotide Polymorphisms of Inflammation-Related Factors Associated with Risk of Lung Cancer." *Medical Oncology (Northwood, London, England)* 30, no. 1 (March 2013): 414. doi:10.1007/s12032-012-0414-6.

Balkwill, Fran, and Alberto Mantovani. "Inflammation and Cancer: Back to Virchow?" *The Lancet* 357, no. 9255 (2001): 539–45.

Bandres, Eva, Nerea Bitarte, Fernando Arias, Jackeline Agorreta, Puri Fortes, Xabi Agirre, Ruth Zarate, et al. "microRNA-451 Regulates Macrophage Migration Inhibitory Factor Production and Proliferation of Gastrointestinal Cancer Cells." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 15, no. 7 (April 1, 2009): 2281–90. doi:10.1158/1078-0432.CCR-08-1818.

Bartel, David P., and Chang-Zheng Chen. "Micromanagers of Gene Expression: The Potentially Widespread Influence of Metazoan microRNAs." *Nature Reviews Genetics* 5, no. 5 (2004): 396–400.

Bartley, A. N., H. Yao, B. A. Barkoh, C. Ivan, B. M. Mishra, A. Rashid, G. A. Calin, R. Luthra, and S. R. Hamilton. "Complex Patterns of Altered MicroRNA Expression during the Adenoma-Adenocarcinoma Sequence for Microsatellite-Stable Colorectal Cancer." *Clinical Cancer Research* 17, no. 23 (December 1, 2011): 7283–93. doi:10.1158/1078-0432.CCR-11-1452.

Beets-Tan, Regina G. H., and Geerard L. Beets. "MRI for Assessing and Predicting Response to Neoadjuvant Treatment in Rectal Cancer." *Nature Reviews Gastroenterology & Hepatology* 11, no. 8 (March 25, 2014): 480–88. doi:10.1038/nrgastro.2014.41.

Benson, Al B. III, Venook, Alan P., Bekaai-Saab Tenios, Chan Emily, Chen Yi-Jen,Cooper Harry S., Engstrom P.F., et al. "Rectal cancer, Version 2.2015 Featured Updates to the NCCN Guidelines." *Journal of the National Comprehensive Cancer Network: JNCCN,*13 no.6 (June 2015): 719-728.

Bluth, Martin, and Jing Li. "Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Implications for Cancer Therapy." *Pharmacogenomics and Personalized Medicine*, April 2011, 11. doi:10.2147/PGPM.S18861.

Boni, V., R. Zarate, J. C. Villa, E. Bandrés, M. A. Gomez, E. Maiello, J. Garcia-Foncillas, and E. Aranda. "Role of Primary miRNA Polymorphic Variants in Metastatic Colon Cancer Patients Treated with 5-Fluorouracil and Irinotecan." *The Pharmacogenomics Journal* 11, no. 6 (December 2011): 429–36. doi:10.1038/tpj.2010.58.

Bouvy, Jacoline C., Marie L. De Bruin, and Marc A. Koopmanschap. "Epidemiology of Adverse Drug Reactions in Europe: A Review of Recent Observational Studies." *Drug Safety* 38, no. 5 (May 2015): 437–53. doi:10.1007/s40264-015-0281-0.

Brandis, Katrina A., Sarah Gale, Sarah Jinn, Stephen J. Langmade, Nicole Dudley-Rucker, Hui Jiang, Rohini Sidhu, et al. "Box C/D Small Nucleolar RNA (snoRNA) U60 Regulates Intracellular Cholesterol Trafficking." *The Journal of Biological Chemistry* 288, no. 50 (December 13, 2013): 35703–13. doi:10.1074/jbc.M113.488577.

Brettingham-Moore, K. H., C. P. Duong, D. M. Greenawalt, A. G. Heriot, J. Ellul, C. A. Dow, W. K. Murray, et al. "Pretreatment Transcriptional Profiling for Predicting Response to Neoadjuvant Chemoradiotherapy in Rectal Adenocarcinoma." *Clinical Cancer Research* 17, no. 9 (May 1, 2011): 3039–47. doi:10.1158/1078-0432.CCR-10-2915.

Cai, Xuezhong, Curt H. Hagedorn, and Bryan R. Cullen. "Human microRNAs Are Processed from Capped, Polyadenylated Transcripts That Can Also Function as mRNAs." *RNA (New York, N.Y.)* 10, no. 12 (December 2004): 1957–66. doi:10.1261/rna.7135204.

Calin, George Adrian, Manuela Ferracin, Amelia Cimmino, Gianpiero Di Leva, Masayoshi Shimizu, Sylwia E. Wojcik, Marilena V. Iorio, et al. "A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia." *The New England Journal of Medicine* 353, no. 17 (October 27, 2005): 1793–1801. doi:10.1056/NEJMoa050995.

Cappuzzo, Federico, Andrea Sacconi, Lorenza Landi, Vienna Ludovini, Francesca Biagioni, Armida D'Incecco, Alessandra Capodanno, et al. "MicroRNA Signature in Metastatic Colorectal Cancer Patients Treated with Anti-EGFR Monoclonal Antibodies." *Clinical Colorectal Cancer* 13, no. 1 (March 2014): 37–45.e4. doi:10.1016/j.clcc.2013.11.006.

Castanotto, Daniela, Robert Lingeman, Arthur D. Riggs, and John J. Rossi. "CRM1 Mediates Nuclear-Cytoplasmic Shuttling of Mature microRNAs." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 51 (December 22, 2009): 21655–59. doi:10.1073/pnas.0912384106.

Cecchin, E., M. Agostini, S. Pucciarelli, A. De Paoli, V. Canzonieri, R. Sigon, E. De Mattia, et al. "Tumor Response Is Predicted by Patient Genetic Profile in Rectal Cancer Patients Treated with Neo-Adjuvant Chemo-Radiotherapy." *The Pharmacogenomics Journal* 11, no. 3 (June 2011): 214–26. doi:10.1038/tpj.2010.25.

Chang, Daniel T., Rish K. Pai, Lisa A. Rybicki, Michael A. Dimaio, Maneesha Limaye, Priya Jayachandran, Albert C. Koong, et al. "Clinicopathologic and Molecular Features of Sporadic Early-Onset Colorectal Adenocarcinoma: An Adenocarcinoma with Frequent Signet Ring Cell Differentiation, Rectal and Sigmoid Involvement, and Adverse Morphologic Features." *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc* 25, no. 8 (August 2012): 1128–39. doi:10.1038/modpathol.2012.61.

Cheng, Jason Chia-Hsien, Madeline S. Graber, Feng-Ming Hsu, Chiao-Ling Tsai, Leon Castaneda, Jang-Ming Lee, Daniel T. Chang, and Albert C. Koong. "High Serum Levels

of Vascular Endothelial Growth Factor-A and Transforming Growth Factor-β1 Before Neoadjuvant Chemoradiotherapy Predict Poor Outcomes in Patients with Esophageal Squamous Cell Carcinoma Receiving Combined Modality Therapy." *Annals of Surgical Oncology* 21, no. 7 (July 2014): 2361–68. doi:10.1245/s10434-014-3611-z.

Cheng, Jia, Hongxia Deng, Bingxiu Xiao, Hui Zhou, Fei Zhou, Zhisen Shen, and Junming Guo. "piR-823, a Novel Non-Coding Small RNA, Demonstrates in Vitro and in Vivo Tumor Suppressive Activity in Human Gastric Cancer Cells." *Cancer Letters* 315, no. 1 (February 1, 2012): 12–17. doi:10.1016/j.canlet.2011.10.004.

Chin, Lena J., Elena Ratner, Shuguang Leng, Rihong Zhai, Sunitha Nallur, Imran Babar, Roman-Ulrich Muller, et al. "A SNP in a Let-7 microRNA Complementary Site in the KRAS 3' Untranslated Region Increases Non-Small Cell Lung Cancer Risk." *Cancer Research* 68, no. 20 (October 15, 2008): 8535–40. doi:10.1158/0008-5472.CAN-08-2129.

Chiosea, Simion, Elena Jelezcova, Uma Chandran, Marie Acquafondata, Teresa McHale, Robert W. Sobol, and Rajiv Dhir. "Up-Regulation of Dicer, a Component of the MicroRNA Machinery, in Prostate Adenocarcinoma." *The American Journal of Pathology* 169, no. 5 (November 2006): 1812–20. doi:10.2353/ajpath.2006.060480.

Chiyomaru, Takeshi, Shinichiro Fukuhara, Sharanjot Saini, Shahana Majid, Guoren Deng, Varahram Shahryari, Inik Chang, et al. "Long Non-Coding RNA HOTAIR Is Targeted and Regulated by miR-141 in Human Cancer Cells." *The Journal of Biological Chemistry* 289, no. 18 (May 2, 2014): 12550–65. doi:10.1074/jbc.M113.488593.

Chu, Haiyan, Liping Xia, Xiaonan Qiu, Dongying Gu, Linjun Zhu, Jing Jin, Gaoyun Hui, et al. "Genetic Variants in Noncoding PIWI-Interacting RNA and Colorectal Cancer Risk: piRNA SNPs and Colorectal Cancer Risk." *Cancer* 121, no. 12 (June 15, 2015): 2044–52. doi:10.1002/cncr.29314.

Chu, Liang, Mack Y. Su, Leonard B. Maggi, Lan Lu, Chelsea Mullins, Seth Crosby, Gaofeng Huang, Wee Joo Chng, Ravi Vij, and Michael H. Tomasson. "Multiple Myeloma–associated Chromosomal Translocation Activates Orphan snoRNA ACA11 to Suppress Oxidative Stress." *The Journal of Clinical Investigation* 122, no. 8 (2012): 2793–2806.

Chung, Arthur C. K., Xiao R. Huang, Xiaoming Meng, and Hui Y. Lan. "miR-192 Mediates TGF-beta/Smad3-Driven Renal Fibrosis." *Journal of the American Society of Nephrology: JASN* 21, no. 8 (August 2010): 1317–25. doi:10.1681/ASN.2010020134.

Chung, Yuan-Chiang, Yuan-Liang Chaen, and Chi-Pin Hsu. "Clinical Significance of Tissue Expression of Interleukin-6 in Colorectal Carcinoma." *Anticancer Research* 26, no. 5B (2006): 3905–11.

Cohen, Eran, Dana Avrahami, Kati Frid, Tamar Canello, Ephrat Levy Lahad, Sharon Zeligson, Shira Perlberg, et al. "Snord 3A: A Molecular Marker and Modulator of Prion Disease Progression." *PloS One* 8, no. 1 (2013): e54433. doi:10.1371/journal.pone.0054433.

Colotta, Francesco, Paola Allavena, Antonio Sica, Cecilia Garlanda, and Alberto Mantovani. "Cancer-Related Inflammation, the Seventh Hallmark of Cancer: Links to Genetic Instability." *Carcinogenesis* 30, no. 7 (July 2009): 1073–81. doi:10.1093/carcin/bgp127.

Cross-Knorr, Sam, Shaolei Lu, Kimberly Perez, Sara Guevara, Kate Brilliant, Claudio Pisano, Peter J. Quesenberry, Murray B. Resnick, and Devasis Chatterjee. "RKIP Phosphorylation and STAT3 Activation Is Inhibited by Oxaliplatin and Camptothecin

and Are Associated with Poor Prognosis in Stage II Colon Cancer Patients." *BMC Cancer* 13 (2013): 463. doi:10.1186/1471-2407-13-463.

Cua, Daniel J., and Cristina M. Tato. "Innate IL-17-Producing Cells: The Sentinels of the Immune System." *Nature Reviews. Immunology* 10, no. 7 (July 2010): 479–89. doi:10.1038/nri2800.

d'Adda di Fagagna, Fabrizio. "A Direct Role for Small Non-Coding RNAs in DNA Damage Response." *Trends in Cell Biology* 24, no. 3 (March 2014): 171–78. doi:10.1016/j.tcb.2013.09.008.

Davies, Janine M., Dimitri Trembath, Allison M. Deal, William K. Funkhouser, Benjamin F. Calvo, Timothy Finnegan, Karen E. Weck, Joel E. Tepper, and Bert H. O'Neil. "Phospho-ERK and AKT Status, but Not KRAS Mutation Status, Are Associated with Outcomes in Rectal Cancer Treated with Chemoradiotherapy." *Radiation Oncology* 6, no. 1 (2011): 1.

Davis, Brandi N., Aaron C. Hilyard, Giorgio Lagna, and Akiko Hata. "SMAD Proteins Control DROSHA-Mediated microRNA Maturation." *Nature* 454, no. 7200 (July 3, 2008): 56–61. doi:10.1038/nature07086.

Davis, Brandi N., Aaron C. Hilyard, Peter H. Nguyen, Giorgio Lagna, and Akiko Hata. "Smad Proteins Bind a Conserved RNA Sequence to Promote microRNA Maturation by Drosha." *Molecular Cell* 39, no. 3 (August 13, 2010): 373–84. doi:10.1016/j.molcel.2010.07.011.

Debucquoy, Annelies, Karin Haustermans, Anneleen Daemen, Selda Aydin, Louis Libbrecht, Olivier Gevaert, Bart De Moor, et al. "Molecular Response to Cetuximab and Efficacy of Preoperative Cetuximab-Based Chemoradiation in Rectal Cancer." *Journal of Clinical Oncology* 27, no. 17 (2009): 2751–57.

De Martel, Catherine, Jacques Ferlay, Silvia Franceschi, Jérôme Vignat, Freddie Bray, David Forman, and Martyn Plummer. "Global Burden of Cancers Attributable to Infections in 2008: A Review and Synthetic Analysis." *The Lancet Oncology* 13, no. 6 (2012): 607–15.

De Paoli, A., S. Chiara, G. Luppi, M. L. Friso, G. D. Beretta, S. Del Prete, L. Pasetto, et al. "Capecitabine in Combination with Preoperative Radiation Therapy in Locally Advanced, Resectable, Rectal Cancer: A Multicentric Phase II Study." *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 17, no. 2 (February 2006): 246–51. doi:10.1093/annonc/mdj041.

Dong, Zhuo, Zhiwei Zhong, Lihua Yang, Shaomin Wang, and Zhaohui Gong. "MicroRNA-31 Inhibits Cisplatin-Induced Apoptosis in Non-Small Cell Lung Cancer Cells by Regulating the Drug Transporter ABCB9." *Cancer Letters* 343, no. 2 (February 2014): 249–57. doi:10.1016/j.canlet.2013.09.034.

Dreussi, Eva, Paola Biason, Giuseppe Toffoli, and Erika Cecchin. "miRNA Pharmacogenomics: The New Frontier for Personalized Medicine in Cancer?" *Pharmacogenomics* 13, no. 14 (November 2012): 1635–50. doi:10.2217/pgs.12.147.

Eng, L., A. K. Azad, S. Habbous, V. Pang, W. Xu, A. H. Maitland-van der Zee, S. Savas, H. J. Mackay, E. Amir, and G. Liu. "Vascular Endothelial Growth Factor Pathway Polymorphisms as Prognostic and Pharmacogenetic Factors in Cancer: A Systematic Review and Meta-Analysis." *Clinical Cancer Research* 18, no. 17 (September 1, 2012): 4526–37. doi:10.1158/1078-0432.CCR-12-1315.

FDA. Paving the Way for Personalized Medicine. (2013).

Foster, Jake D., Emma L. Jones, Stephen Falk, Edwin J. Cooper, and Nader K. Francis. "Timing of Surgery after Long-Course Neoadjuvant Chemoradiotherapy for Rectal Cancer: A Systematic Review of the Literature." *Diseases of the Colon and Rectum* 56, no. 7 (July 2013): 921–30. doi:10.1097/DCR.0b013e31828aedcb.

Gabrilovich, Dmitry I., and Srinivas Nagaraj. "Myeloid-Derived Suppressor Cells as Regulators of the Immune System." *Nature Reviews. Immunology* 9, no. 3 (March 2009): 162–74. doi:10.1038/nri2506.

Gagnon, Keith T., and David R. Corey. "Argonaute and the Nuclear RNAs: New Pathways for RNA-Mediated Control of Gene Expression." *Nucleic Acid Therapeutics* 22, no. 1 (February 2012): 3–16. doi:10.1089/nat.2011.0330.

Gal, Hilah, Gopal Pandi, Andrew A. Kanner, Zvi Ram, Gila Lithwick-Yanai, Ninette Amariglio, Gideon Rechavi, and David Givol. "MIR-451 and Imatinib Mesylate Inhibit Tumor Growth of Glioblastoma Stem Cells." *Biochemical and Biophysical Research Communications* 376, no. 1 (November 7, 2008): 86–90. doi:10.1016/j.bbrc.2008.08.107.

Galon, Jérôme, Helen K. Angell, Davide Bedognetti, and Francesco M. Marincola. "The Continuum of Cancer Immunosurveillance: Prognostic, Predictive, and Mechanistic Signatures." *Immunity* 39, no. 1 (July 2013): 11–26. doi:10.1016/j.immuni.2013.07.008.

Galon, Jérôme, Franck Pagès, Francesco M. Marincola, Helen K. Angell, Magdalena Thurin, Alessandro Lugli, Inti Zlobec, et al. "Cancer Classification Using the Immunoscore: A Worldwide Task Force." *Journal of Translational Medicine* 10 (2012): 205. doi:10.1186/1479-5876-10-205.

Gomes dos Santos, L., Lages, R.B., Modesto Alves, G.C., Araùjo,and J.K.L. "Assessment of tumor regression in patients with rectal carcinoma treated with neoadjuvat chemoradiotherapy". *Applied Cancer Research* 32(3) (October 2012):64-69.

Ganzinelli, Monica, Eliana Rulli, Elisa Caiola, Marina Chiara Garassino, Massimo Broggini, Elena Copreni, Sheila Piva, et al. "Role of KRAS-LCS6 Polymorphism in Advanced NSCLC Patients Treated with Erlotinib or Docetaxel in Second Line Treatment (TAILOR)." *Scientific Reports* 5 (November 17, 2015): 16331. doi:10.1038/srep16331.

García-Flórez, Luis Joaquín, Guillermo Gómez-Álvarez, Ana Madalina Frunza, Luis Barneo-Serra, Carmen Martínez-Alonso, and Manuel Florentino Fresno-Forcelledo. "Predictive Markers of Response to Neoadjuvant Therapy in Rectal Cancer." *Journal of Surgical Research* 194, no. 1 (March 2015): 120–26. doi:10.1016/j.jss.2014.10.005.

Glimelius, B., Tiret, E., Cervantes, A., Arnold D, and ESMO Guidelines Working Group. "Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." *Ann Oncol*. (Oct 2013);24 Suppl 6:vi81-8. doi: 10.1093/annonc/mdt240.

Gopinath, S. C. B., A. Matsugami, M. Katahira, and P. K. R. Kumar. "Human Vault-Associated Non-Coding RNAs Bind to Mitoxantrone, a Chemotherapeutic Compound." *Nucleic Acids Research* 33, no. 15 (September 8, 2005): 4874–81. doi:10.1093/nar/gki809.

Gravitz, Lauren. "Therapy: This Time It's Personal." *Nature* 509, no. 7502 (May 29, 2014): S52–54. doi:10.1038/509S52a.

Grelier, G., N. Voirin, A.-S. Ay, D. G. Cox, S. Chabaud, I. Treilleux, S. Léon-Goddard, et al. "Prognostic Value of Dicer Expression in Human Breast Cancers and Association

with the Mesenchymal Phenotype." *British Journal of Cancer* 101, no. 4 (August 18, 2009): 673–83. doi:10.1038/sj.bjc.6605193.

Grivennikov, Sergei I., Florian R. Greten, and Michael Karin. "Immunity, Inflammation, and Cancer." *Cell* 140, no. 6 (March 2010): 883–99. doi:10.1016/j.cell.2010.01.025.

Grivna, Shane T., Ergin Beyret, Zhong Wang, and Haifan Lin. "A Novel Class of Small RNAs in Mouse Spermatogenic Cells." *Genes & Development* 20, no. 13 (July 1, 2006): 1709–14. doi:10.1101/gad.1434406.

Grulich, Andrew E., Marina T. van Leeuwen, Michael O. Falster, and Claire M. Vajdic. "Incidence of Cancers in People with HIV/AIDS Compared with Immunosuppressed Transplant Recipients: A Meta-Analysis." *Lancet (London, England)* 370, no. 9581 (July 7, 2007): 59–67. doi:10.1016/S0140-6736(07)61050-2.

Guthrie, Graeme J. K., Campbell S. D. Roxburgh, Paul G. Horgan, and Donald C. McMillan. "Does Interleukin-6 Link Explain the Link between Tumour Necrosis, Local and Systemic Inflammatory Responses and Outcome in Patients with Colorectal Cancer?" *Cancer Treatment Reviews* 39, no. 1 (February 2013): 89–96. doi:10.1016/j.ctrv.2012.07.003.

Hall, J. S., J. Taylor, H. R. Valentine, J. J. Irlam, A. Eustace, P. J. Hoskin, C. J. Miller, and C. M. L. West. "Enhanced Stability of microRNA Expression Facilitates Classification of FFPE Tumour Samples Exhibiting near Total mRNA Degradation." *British Journal of Cancer* 107, no. 4 (August 7, 2012): 684–94. doi:10.1038/bjc.2012.294.

Hanahan, Douglas, and Robert A. Weinberg. "Hallmarks of Cancer: The next Generation." *Cell* 144, no. 5 (March 4, 2011): 646–74. doi:10.1016/j.cell.2011.02.013.

Huang, Ming-Yii, Chih-Hung Lin, Chun-Ming Huang, Hsiang-Lin Tsai, Ching-Wen Huang, Yung-Sung Yeh, Chee-Yin Chai, and Jaw-Yuan Wang. "Relationships Between SMAD3 Expression and Preoperative Fluoropyrimidine-Based Chemoradiotherapy Response in Locally Advanced Rectal Cancer Patients." *World Journal of Surgery* 39, no. 5 (May 2015): 1257–67. doi:10.1007/s00268-014-2917-0.

Huang, Vera, and Long-Cheng Li. "miRNA Goes Nuclear." *RNA Biology* 9, no. 3 (March 2012): 269–73. doi:10.4161/rna.19354.

Hubackova, S, A Kucerova, G Michlits, L Kyjacova, M Reinis, O Korolov, J Bartek, and Z Hodny. "IFNγ Induces Oxidative Stress, DNA Damage and Tumor Cell Senescence via TGFβ/SMAD Signaling-Dependent Induction of Nox4 and Suppression of ANT2." *Oncogene*, May 18, 2015. doi:10.1038/onc.2015.162.

Huerta, Sergio, Diane M. Harris, Ali Jazirehi, Benjamin Bonavida, David Elashoff, Edward H. Livingston, and David Heber. "Gene Expression Profile of Metastatic Colon Cancer Cells Resistant to Cisplatin-Induced Apoptosis." *International Journal of Oncology* 22, no. 3 (March 2003): 663–70.

Hur, Hyuk, Nam Kyu Kim, Byung Soh Min, Seung Hyuk Baik, Kang Young Lee, Woong Sub Koom, Joong Bae Ahn, and Hoguen Kim. "Can a Biomarker-Based Scoring System Predict Pathologic Complete Response After Preoperative Chemoradiotherapy for Rectal Cancer?:" *Diseases of the Colon & Rectum* 57, no. 5 (May 2014): 592–601. doi:10.1097/DCR.0000000000000109.

Ishigame, Harumichi, Shigeru Kakuta, Takeshi Nagai, Motohiko Kadoki, Aya Nambu, Yutaka Komiyama, Noriyuki Fujikado, et al. "Differential Roles of Interleukin-17A and -17F in Host Defense against Mucoepithelial Bacterial Infection and Allergic Responses." *Immunity* 30, no. 1 (January 16, 2009): 108–19. doi:10.1016/j.immuni.2008.11.009.

Izquierdo, M. A., G. L. Scheffer, A. B. Schroeijers, M. C. de Jong, and R. J. Scheper. "Vault-Related Resistance to Anticancer Drugs Determined by the Expression of the Major Vault Protein LRP." *Cytotechnology* 27, no. 1–3 (September 1998): 137–48. doi:10.1023/A:1008004502861.

Jakymiw, Andrew, Rushi S. Patel, Natasha Deming, Indraneel Bhattacharyya, Priya Shah, Richard J. Lamont, Carol M. Stewart, Donald M. Cohen, and Edward K. L. Chan. "Overexpression of Dicer as a Result of Reduced Let-7 MicroRNA Levels Contributes to Increased Cell Proliferation of Oral Cancer Cells." *Genes, Chromosomes & Cancer* 49, no. 6 (June 2010): 549–59. doi:10.1002/gcc.20765.

Jiang, Y., P. J. Villeneuve, A. Wielgosz, D. E. Schaubel, S. S. A. Fenton, and Y. Mao. "The Incidence of Cancer in a Population-Based Cohort of Canadian Heart Transplant Recipients." *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 10, no. 3 (March 2010): 637–45. doi:10.1111/j.1600-6143.2009.02973.x.

Ji, Juling, Jinsheng Zhang, Guangcun Huang, Jin Qian, Xueqing Wang, and Shuang Mei. "Over-Expressed microRNA-27a and 27b Influence Fat Accumulation and Cell Proliferation during Rat Hepatic Stellate Cell Activation." *FEBS Letters* 583, no. 4 (February 18, 2009): 759–66. doi:10.1016/j.febslet.2009.01.034.

Kalbasi, Anusha, Carl H. June, Naomi Haas, and Neha Vapiwala. "Radiation and Immunotherapy: A Synergistic Combination." *Journal of Clinical Investigation* 123, no. 7 (July 1, 2013): 2756–63. doi:10.1172/JCI69219.

Karube, Yoko, Hisaaki Tanaka, Hirotaka Osada, Shuta Tomida, Yoshio Tatematsu, Kiyoshi Yanagisawa, Yasushi Yatabe, et al. "Reduced Expression of Dicer Associated with Poor Prognosis in Lung Cancer Patients." *Cancer Science* 96, no. 2 (February 2005): 111–15. doi:10.1111/j.1349-7006.2005.00015.x.

Kawai, S., and A. Amano. "BRCA1 Regulates microRNA Biogenesis via the DROSHA Microprocessor Complex." *The Journal of Cell Biology* 197, no. 2 (April 16, 2012): 201–8. doi:10.1083/jcb.201110008.

Kedersha, N. L., J. E. Heuser, D. C. Chugani, and L. H. Rome. "Vaults. III. Vault Ribonucleoprotein Particles Open into Flower-like Structures with Octagonal Symmetry." *The Journal of Cell Biology* 112, no. 2 (January 1991): 225–35.

Kjersem, Janne B., Tone Ikdahl, Tormod Guren, Eva Skovlund, Halfdan Sorbye, Julian Hamfjord, Per Pfeiffer, et al. "Let-7 miRNA-Binding Site Polymorphism in the KRAS 3′ UTR; Colorectal Cancer Screening Population Prevalence and Influence on Clinical Outcome in Patients with Metastatic Colorectal Cancer Treated with 5-Fluorouracil and Oxaliplatin+/- Cetuximab." *BMC Cancer* 12, no. 1 (2012): 534.

Kogo, Ryunosuke, Teppei Shimamura, Koshi Mimori, Kohichi Kawahara, Seiya Imoto, Tomoya Sudo, Fumiaki Tanaka, et al. "Long Noncoding RNA HOTAIR Regulates Polycomb-Dependent Chromatin Modification and Is Associated with Poor Prognosis in Colorectal Cancers." *Cancer Research* 71, no. 20 (October 15, 2011): 6320–26. doi:10.1158/0008-5472.CAN-11-1021.

Kong, Lu, Qi Hao, Ying Wang, Ping Zhou, Binbin Zou, and Yu-xiang Zhang. "Regulation of p53 Expression and Apoptosis by Vault RNA2-1-5p in Cervical Cancer Cells." *Oncotarget* 6, no. 29 (2015): 28371.

Krol, Jacek, Inga Loedige, and Witold Filipowicz. "The Widespread Regulation of microRNA Biogenesis, Function and Decay." *Nature Reviews Genetics*, July 27, 2010. doi:10.1038/nrg2843.

Kuehbacher, Angelika, Carmen Urbich, Andreas M. Zeiher, and Stefanie Dimmeler. "Role of Dicer and Drosha for Endothelial microRNA Expression and Angiogenesis." *Circulation Research* 101, no. 1 (July 6, 2007): 59–68. doi:10.1161/CIRCRESAHA.107.153916.

Kulda, Vlastimil, Martin Pesta, Ondrej Topolcan, Vaclav Liska, Vladislav Treska, Alan Sutnar, Karel Rupert, et al. "Relevance of miR-21 and miR-143 Expression in Tissue Samples of Colorectal Carcinoma and Its Liver Metastases." *Cancer Genetics and Cytogenetics* 200, no. 2 (July 15, 2010): 154–60. doi:10.1016/j.cancergencyto.2010.04.015.

Lamba, V., Y. Ghodke, W. Guan, and T.S. Tracy. "microRNA-34a Is Associated with Expression of Key Hepatic Transcription Factors and Cytochromes P450." *Biochemical and Biophysical Research Communications* 445, no. 2 (March 2014): 404–11. doi:10.1016/j.bbrc.2014.02.024.

Landi, Stefano, Fabio Bottari, Federica Gemignani, Lydie Gioia-Patricola, Elisabeth Guino, Alfonso Osorio, Javier de Oca, et al. "Interleukin-4 and Interleukin-4 Receptor Polymorphisms and Colorectal Cancer Risk." *European Journal of Cancer (Oxford, England: 1990)* 43, no. 4 (March 2007): 762–68. doi:10.1016/j.ejca.2006.10.024.

Landi, Stefano, Victor Moreno, Lydie Gioia-Patricola, Elisabeth Guino, Matilde Navarro, Javier de Oca, Gabriel Capella, Federico Canzian, and Bellvitge Colorectal Cancer Study Group. "Association of Common Polymorphisms in Inflammatory Genes Interleukin (IL)6, IL8, Tumor Necrosis Factor Alpha, NFKB1, and Peroxisome Proliferator-Activated Receptor Gamma with Colorectal Cancer." *Cancer Research* 63, no. 13 (July 1, 2003): 3560–66.

Law, Priscilla T.-Y., Hao Qin, Arthur K.-K. Ching, Keng Po Lai, Ngai Na Co, Mian He, Raymond W.-M. Lung, Anthony W.-H. Chan, Ting-Fung Chan, and Nathalie Wong. "Deep Sequencing of Small RNA Transcriptome Reveals Novel Non-Coding RNAs in Hepatocellular Carcinoma." *Journal of Hepatology* 58, no. 6 (June 2013): 1165–73. doi:10.1016/j.jhep.2013.01.032.

Lee, Ho-June, Guanglei Zhuang, Yi Cao, Pan Du, Hyo-Jin Kim, and Jeff Settleman. "Drug Resistance via Feedback Activation of Stat3 in Oncogene-Addicted Cancer Cells." *Cancer Cell* 26, no. 2 (August 11, 2014): 207–21. doi:10.1016/j.ccr.2014.05.019.

Liao, John B. "Viruses and Human Cancer." *Yale J Biol Med* 79, no. 3–4 (2006): 115–22.

Li, Cuiping, Bingshuang Long, Xianjing Qin, Weixiong Li, and Yang Zhou. "Cytochrome P1B1 (CYP1B1) Polymorphisms and Cancer Risk: A Meta-Analysis of 52 Studies." *Toxicology* 327 (January 2, 2015): 77–86. doi:10.1016/j.tox.2014.11.007.

Li, Jing, and Martin H Bluth. "Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Implications for Cancer Therapy." *Pharmacogenomics and Personalized Medicine* 4 (April 1, 2011): 11–33. doi:10.2147/PGPM.S18861.

Li, Jing-Hua, Xue Xiao, Yao-Nan Zhang, Ya-Mei Wang, Li-Min Feng, Yu-Mei Wu, and Yu-Xiang Zhang. "MicroRNA miR-886-5p Inhibits Apoptosis by down-Regulating Bax Expression in Human Cervical Carcinoma Cells." *Gynecologic Oncology* 120, no. 1 (January 2011): 145–51. doi:10.1016/j.ygyno.2010.09.009.

Li, Long-Cheng, Steven T. Okino, Hong Zhao, Deepa Pookot, Robert F. Place, Shinji Urakami, Hideki Enokida, and Rajvir Dahiya. "Small dsRNAs Induce Transcriptional Activation in Human Cells." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 46 (November 14, 2006): 17337–42. doi:10.1073/pnas.0607015103.

Li, Ning, Chao Zhang, Zhaoquan Chen, Lilu Bai, Min Nie, Bin Zhou, and Huanxi Xu. "Interleukin 17A and Interleukin 17F Polymorphisms Are Associated with Oral Squamous Cell Carcinoma Susceptibility in a Chinese Population." *Journal of Oral and Maxillofacial Surgery: Official Journal of the American Association of Oral and Maxillofacial Surgeons* 73, no. 2 (February 2015): 267–73. doi:10.1016/j.joms.2014.09.009.

Liu, Jidong, Marco Antonio Valencia-Sanchez, Gregory J. Hannon, and Roy Parker. "MicroRNA-Dependent Localization of Targeted mRNAs to Mammalian P-Bodies." *Nature Cell Biology* 7, no. 7 (July 2005): 719–23. doi:10.1038/ncb1274.

Li, Xin, Yu-Zhuo Pan, Gail M. Seigel, Zi-Hua Hu, Min Huang, and Ai-Ming Yu. "Breast Cancer Resistance Protein BCRP/ABCG2 Regulatory microRNAs (hsa-miR-328, -519c and -520h) and Their Differential Expression in Stem-like ABCG2+ Cancer Cells." *Biochemical Pharmacology* 81, no. 6 (March 15, 2011): 783–92. doi:10.1016/j.bcp.2010.12.018.

Li, X., Y. Liu, K. J. Granberg, Q. Wang, L. M. Moore, P. Ji, J. Gumin, et al. "Two Mature Products of MIR-491 Coordinate to Suppress Key Cancer Hallmarks in Glioblastoma." *Oncogene* 34, no. 13 (March 26, 2015): 1619–28. doi:10.1038/onc.2014.98.

Luby, Thomas M., Geoffrey Cole, Lisa Baker, J. Steven Kornher, Urban Ramstedt, and Mary Lynne Hedley. "Repeated Immunization with Plasmid DNA Formulated in Poly(lactide-Co-Glycolide) Microparticles Is Well Tolerated and Stimulates Durable T Cell Responses to the Tumor-Associated Antigen Cytochrome P450 1B1." *Clinical Immunology (Orlando, Fla.)* 112, no. 1 (July 2004): 45–53. doi:10.1016/j.clim.2004.04.002.

Lu, Jun, Gad Getz, Eric A. Miska, Ezequiel Alvarez-Saavedra, Justin Lamb, David Peck, Alejandro Sweet-Cordero, et al. "MicroRNA Expression Profiles Classify Human Cancers." *Nature* 435, no. 7043 (June 9, 2005): 834–38. doi:10.1038/nature03702.

Lukowski, Samuel W., Richard J. Fish, Juliette Martin-Levilain, Carmen Gonelle-Gispert, Leo H. Bühler, Pierre Maechler, Emmanouil T. Dermitzakis, and Marguerite Neerman-Arbez. "Integrated Analysis of mRNA and miRNA Expression in Response to Interleukin-6 in Hepatocytes." *Genomics* 106, no. 2 (Agosto 2015): 107–15. doi:10.1016/j.ygeno.2015.05.001.

Maas, M., R. G. H. Beets-Tan, D. M. J. Lambregts, G. Lammering, P. J. Nelemans, S. M. E. Engelen, R. M. van Dam, et al. "Wait-and-See Policy for Clinical Complete Responders After Chemoradiation for Rectal Cancer." *Journal of Clinical Oncology* 29, no. 35 (December 10, 2011): 4633–40. doi:10.1200/JCO.2011.37.7176.

MacFarlane, Leigh-Ann, and Paul R. Murphy. "MicroRNA: Biogenesis, Function and Role in Cancer." *Current Genomics* 11, no. 7 (November 2010): 537–61. doi:10.2174/138920210793175895.

Macleod, K. F., N. Sherry, G. Hannon, D. Beach, T. Tokino, K. Kinzler, B. Vogelstein, and T. Jacks. "p53-Dependent and Independent Expression of p21 during Cell Growth, Differentiation, and DNA Damage." *Genes & Development* 9, no. 8 (April 15, 1995): 935–44.

Maher, Brendan. "ENCODE: The Human Encyclopaedia." *Nature* 489, no. 7414 (September 6, 2012): 46–48.

Mandard, A. M., F. Dalibard, J. C. Mandard, J. Marnay, M. Henry-Amar, J. F. Petiot, A. Roussel, J. H. Jacob, P. Segol, and G. Samama. "Pathologic Assessment of Tumor Regression after

Preoperative Chemoradiotherapy of Esophageal Carcinoma. Clinicopathologic Correlations." *Cancer* 73, no. 11 (June 1, 1994): 2680–86.

Markkula, Andrea, Maria Simonsson, Christian Ingvar, Carsten Rose, and Helena Jernström. "IL6 Genotype, Tumour ER-Status, and Treatment Predicted Disease-Free Survival in a Prospective Breast Cancer Cohort." *BMC Cancer* 14 (2014): 759. doi:10.1186/1471-2407-14-759.

Martens, Milou H., Samina Subhani, Luc A. Heijnen, Doenja M.J. Lambregts, Jeroen Buijsen, Monique Maas, Robert G. Riedl, et al. "Can Perfusion MRI Predict Response to Preoperative Treatment in Rectal Cancer?" *Radiotherapy and Oncology* 114, no. 2 (February 2015): 218–23. doi:10.1016/j.radonc.2014.11.044.

Martens-Uzunova, Elena S., Michael Olvedy, and Guido Jenster. "Beyond microRNA--Novel RNAs Derived from Small Non-Coding RNA and Their Implication in Cancer." *Cancer Letters* 340, no. 2 (November 1, 2013): 201–11. doi:10.1016/j.canlet.2012.11.058.

Ma, Xiao-Pin, Guopeng Yu, Xubo Chen, Qianyi Xiao, Zhuqing Shi, Lu-Yao Zhang, Haitao Chen, et al. "MiR-608 rs4919510 Is Associated with Prognosis of Hepatocellular Carcinoma." *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, January 27, 2016. doi:10.1007/s13277-016-4897-1.

McFadyen, M. C., H. L. McLeod, F. C. Jackson, W. T. Melvin, J. Doehmer, and G. I. Murray. "Cytochrome P450 CYP1B1 Protein Expression: A Novel Mechanism of Anticancer Drug Resistance." *Biochemical Pharmacology* 62, no. 2 (July 15, 2001): 207–12.

Merritt, William M., Yvonne G. Lin, Liz Y. Han, Aparna A. Kamat, Whitney A. Spannuth, Rosemarie Schmandt, Diana Urbauer, et al. "Dicer, Drosha, and Outcomes in Patients with Ovarian Cancer." *The New England Journal of Medicine* 359, no. 25 (December 18, 2008): 2641–50. doi:10.1056/NEJMoa0803785.

Momen-Heravi, F., A. J. Trachtenberg, W. P. Kuo, and Y. S. Cheng. "Genomewide Study of Salivary MicroRNAs for Detection of Oral Cancer." *Journal of Dental Research* 93, no. 7 Suppl (July 1, 2014): 86S – 93S. doi:10.1177/0022034514531018.

Morikawa, Teppei, Yoshifumi Baba, Mai Yamauchi, Aya Kuchiba, Katsuhiko Nosho, Kaori Shima, Noriko Tanaka, et al. "STAT3 Expression, Molecular Features, Inflammation Patterns, and Prognosis in a Database of 724 Colorectal Cancers." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 17, no. 6 (March 15, 2011): 1452–62. doi:10.1158/1078-0432.CCR-10-2694.

Munch, Erika M., R. Alan Harris, Mahmoud Mohammad, Ashley L. Benham, Sasha M. Pejerrey, Lori Showalter, Min Hu, et al. "Transcriptome Profiling of microRNA by Next-Gen Deep Sequencing Reveals Known and Novel miRNA Species in the Lipid Fraction of Human Breast Milk." *PloS One* 8, no. 2 (2013): e50564. doi:10.1371/journal.pone.0050564.

Nguyen, Andrew V., Yuan-Yuan Wu, and Elaine Y. Lin. "STAT3 and Sphingosine-1-Phosphate in Inflammation-Associated Colorectal Cancer." *World Journal of Gastroenterology* 20, no. 30 (August 14, 2014): 10279–87. doi:10.3748/wjg.v20.i30.10279.

Okamura, Katsutomo, Joshua W. Hagen, Hong Duan, David M. Tyler, and Eric C. Lai. "The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila." *Cell* 130, no. 1 (July 13, 2007): 89–100. doi:10.1016/j.cell.2007.06.028.

Omrane, Inés, Imen Medimegh, Olfa Baroudi, Hager Ayari, Walid Bedhiafi, Nejla Stambouli, Marwa Ferchichi, et al. "Involvement of IL17A, IL17F and IL23R Polymorphisms in

Colorectal Cancer Therapy." *PLOS ONE* 10, no. 6 (giu 2015): e0128911. doi:10.1371/journal.pone.0128911.

Pan, Yu-Zhuo, Wenqing Gao, and Ai-Ming Yu. "MicroRNAs Regulate CYP3A4 Expression via Direct and Indirect Targeting." *Drug Metabolism and Disposition: The Biological Fate of Chemicals* 37, no. 10 (October 2009): 2112–17. doi:10.1124/dmd.109.027680.

Park, Sungmi, Jeong-Kook Kim, Chang Joo Oh, Seung Hee Choi, Jae-Han Jeon, and In-Kyu Lee. "Scoparone Interferes with STAT3-Induced Proliferation of Vascular Smooth Muscle Cells." *Experimental & Molecular Medicine* 47 (2015): e145. doi:10.1038/emm.2014.113.

Paroo, Zain, Xuecheng Ye, She Chen, and Qinghua Liu. "Phosphorylation of the Human microRNA-Generating Complex Mediates MAPK/Erk Signaling." *Cell* 139, no. 1 (October 2, 2009): 112–22. doi:10.1016/j.cell.2009.06.044.

Patel, S A A, U Bhambra, M P Charalambous, R M David, R J Edwards, T Lightfoot, A R Boobis, and N J Gooderham. "Interleukin-6 Mediated Upregulation of CYP1B1 and CYP2E1 in Colorectal Cancer Involves DNA Methylation, miR27b and STAT3." *British Journal of Cancer* 111, no. 12 (December 9, 2014): 2287–96. doi:10.1038/bjc.2014.540.

Peter J. Russell. *iGenetics: A Molecular Approach*. 3th ed., 2002.

Petrelli, Fausto, Andrea Coinu, Veronica Lonati, and Sandro Barni. "A Systematic Review and Meta-Analysis of Adjuvant Chemotherapy after Neoadjuvant Treatment and Surgery for Rectal Cancer." *International Journal of Colorectal Disease* 30, no. 4 (April 2015): 447–57. doi:10.1007/s00384-014-2082-9.

Place, Robert F., Long-Cheng Li, Deepa Pookot, Emily J. Noonan, and Rajvir Dahiya. "MicroRNA-373 Induces Expression of Genes with Complementary Promoter Sequences." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 5 (February 5, 2008): 1608–13. doi:10.1073/pnas.0707594105.

Probst, Christian P., Adan Z. Becerra, Christopher T. Aquina, Mohamedtaki A. Tejani, Steven D. Wexner, Julio Garcia-Aguilar, Feza H. Remzi, et al. "Extended Intervals after Neoadjuvant Therapy in Locally Advanced Rectal Cancer: The Key to Improved Tumor Response and Potential Organ Preservation." *Journal of the American College of Surgeons* 221, no. 2 (August 2015): 430–40. doi:10.1016/j.jamcollsurg.2015.04.010.

Qin, Ancheng, Qiang Yu, Yuan Gao, Jifu Tan, Hai Huang, Zhiming Qiao, and Weifeng Qian. "Inhibition of STAT3/cyclinD1 Pathway Promotes Chemotherapeutic Sensitivity of Colorectal Caner." *Biochemical and Biophysical Research Communications* 457, no. 4 (February 20, 2015): 681–87. doi:10.1016/j.bbrc.2015.01.048.

Rébé, C., Végran, F., Berger, H., and Ghiringhelli F. "STAT3 activation: A key factor in tumor immunoescape".*JAKSTAT* (January 1, 2013);2(1):e23010. doi: 10.4161/jkst.23010.

Rehwinkel, Jan, Isabelle Behm-Ansmant, David Gatfield, and Elisa Izaurralde. "A Crucial Role for GW182 and the DCP1:DCP2 Decapping Complex in miRNA-Mediated Gene Silencing." *RNA (New York, N.Y.)* 11, no. 11 (November 2005): 1640–47. doi:10.1261/rna.2191905.

Rochat, B., J. M. Morsman, G. I. Murray, W. D. Figg, and H. L. McLeod. "Human CYP1B1 and Anticancer Agent Metabolism: Mechanism for Tumor-Specific Drug Inactivation?" *The Journal of Pharmacology and Experimental Therapeutics* 296, no. 2 (February 2001): 537–41.

Rokavec, Matjaz, Meryem Gülfem Öner, Huihui Li, Rene Jackstadt, Longchang Jiang, Dmitri Lodygin, Markus Kaller, et al. "IL-6R/STAT3/miR-34a Feedback Loop Promotes EMT-Mediated Colorectal Cancer Invasion and Metastasis." *The Journal of Clinical Investigation* 124, no. 4 (April 2014): 1853–67. doi:10.1172/JCI73531.

Roy, Paromita, Stefano Serra, Erin Kennedy, and Runjan Chetty. "The Prognostic Value of Grade of Regression and Oncocytic Change in Rectal Adenocarcinoma Treated with Neo-Adjuvant Chemoradiotherapy." *Journal of Surgical Oncology* 105, no. 2 (February 2012): 130–34. doi:10.1002/jso.22073.

Ryan, Jennifer E, Satish K Warrier, A Craig Lynch, Robert G Ramsay, Wayne A Phillips, and Alexander G Heriot. "Predicting Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer: A Systemic Review." *Colorectal Disease*, November 2015, n/a – n/a. doi:10.1111/codi.13207.

Salendo, Junius, Melanie Spitzner, Frank Kramer, Xin Zhang, Peter Jo, Hendrik A. Wolff, Julia Kitz, et al. "Identification of a microRNA Expression Signature for Chemoradiosensitivity of Colorectal Cancer Cells, Involving miRNAs-320a, -224, -132 and let7g." *Radiotherapy and Oncology* 108, no. 3 (September 2013): 451–57. doi:10.1016/j.radonc.2013.06.032.

Sanger, F., S. Nicklen, and A. R. Coulson. "DNA Sequencing with Chain-Terminating Inhibitors." *Proceedings of the National Academy of Sciences of the United States of America* 74, no. 12 (December 1977): 5463–67.

Sapre, Nikhil, Matthew K. H. Hong, Geoff Macintyre, Heather Lewis, Adam Kowalczyk, Anthony J. Costello, Niall M. Corcoran, and Christopher M. Hovens. "Curated microRNAs in Urine and Blood Fail to Validate as Predictive Biomarkers for High-Risk Prostate Cancer." *PloS One* 9, no. 4 (2014): e91729. doi:10.1371/journal.pone.0091729.

Schetter, Aaron J., Suet Yi Leung, Jane J. Sohn, Krista A. Zanetti, Elise D. Bowman, Nozomu Yanaihara, Siu Tsan Yuen, et al. "MicroRNA Expression Profiles Associated with Prognosis and Therapeutic Outcome in Colon Adenocarcinoma." *JAMA* 299, no. 4 (January 30, 2008): 425–36. doi:10.1001/jama.299.4.425.

Schoenfeld, J. D., D. N. Margalit, J. L. Kasperzyk, I. M. Shui, J. R. Rider, M. M. Epstein, A. Meisner, et al. "A Single Nucleotide Polymorphism in Inflammatory Gene RNASEL Predicts Outcome after Radiation Therapy for Localized Prostate Cancer." *Clinical Cancer Research* 19, no. 6 (March 15, 2013): 1612–19. doi:10.1158/1078-0432.CCR-12-2718.

Schork, Nicholas J. "Personalized Medicine: Time for One-Person Trials." *Nature* 520, no. 7549 (2015): 609–11.

Sclafani, F., I. Chau, D. Cunningham, C. Peckitt, A. Lampis, J.C. Hahne, C. Braconi, et al. "Prognostic Role of the LCS6 KRAS Variant in Locally Advanced Rectal Cancer: Results of the EXPERT-C Trial." *Annals of Oncology* 26, no. 9 (September 2015): 1936–41. doi:10.1093/annonc/mdv285.

Simmer, Femke, Sabine Venderbosch, Jeroen R. Dijkstra, Elisa M. Vink-Börger, Claudius Faber, Leonie J. Mekenkamp, Miriam Koopman, Anton F. De Haan, Cornelis J. Punt, and Iris D. Nagtegaal. "MicroRNA-143 Is a Putative Predictive Factor for the Response to Fluoropyrimidine-Based Chemotherapy in Patients with Metastatic Colorectal Cancer." *Oncotarget* 6, no. 26 (September 8, 2015): 22996–7. doi:10.18632/oncotarget.4035.

Sissung, T. M. “Pharmacogenetics and Regulation of Human Cytochrome P450 1B1: Implications in Hormone-Mediated Tumor Metabolism and a Novel Target for Therapeutic Intervention.” *Molecular Cancer Research* 4, no. 3 (March 1, 2006): 135–50. doi:10.1158/1541-7786.MCR-05-0101.

Skinner, Heath D., Jeffrey H. Lee, Manoop S. Bhutani, Brian Weston, Wayne Hofstetter, Ritsuko Komaki, Hironori Shiozaki, et al. “A Validated miRNA Profile Predicts Response to Therapy in Esophageal Adenocarcinoma: miRNA Predicts Response to Therapy.” *Cancer* 120, no. 23 (December 1, 2014): 3635–41. doi:10.1002/cncr.28911.

Slaby, O., M. Svoboda, P. Fabian, T. Smerdova, D. Knoflickova, M. Bednarikova, R. Nenutil, and R. Vyzula. “Altered Expression of miR-21, miR-31, miR-143 and miR-145 Is Related to Clinicopathologic Features of Colorectal Cancer.” *Oncology* 72, no. 5–6 (2007): 397–402. doi:10.1159/000113489.

Slattery, Martha L., and Abbie Lundgreen. “The Influence of the CHIEF Pathway on Colorectal Cancer-Specific Mortality.” Edited by Francisco X. Real. *PLoS ONE* 9, no. 12 (December 26, 2014): e116169. doi:10.1371/journal.pone.0116169.

Sohn, Byeong Seok, Seong Joon Park, Jeong Eun Kim, Kyu-pyo Kim, Yong Sang Hong, Cheolwon Suh, Yong Sung Kim, et al. “Single-Nucleotide Polymorphisms in the Vascular Endothelial Growth Factor Pathway and Outcomes of Patients Treated with First-Line Cytotoxic Chemotherapy Combined with Bevacizumab for Advanced Colorectal Cancer.” *Oncology* 87, no. 5 (2014): 280–92. doi:10.1159/000365593.

Song, Buer, Kristine D. Estrada, and Karen M. Lyons. “Smad Signaling in Skeletal Development and Regeneration.” *Cytokine & Growth Factor Reviews* 20, no. 5–6 (2009): 379–88. doi:10.1016/j.cytogfr.2009.10.010.

Spear, Brian B., Margo Heath-Chiozzi, and Jeffrey Huff. “Clinical Application of Pharmacogenetics.” *Trends in Molecular Medicine* 7, no. 5 (2001): 201–4.

Spitzner, Melanie, Birte Roesler, Christian Bielfeld, Georg Emons, Jochen Gaedcke, Hendrik A. Wolff, Margret Rave-Fränk, et al. “STAT3 Inhibition Sensitizes Colorectal Cancer to Chemoradiotherapy in Vitro and in Vivo.” *International Journal of Cancer. Journal International Du Cancer* 134, no. 4 (February 15, 2014): 997–1007. doi:10.1002/ijc.28429.

Strimbu, Kyle, and Jorge A. Tavel. “What Are Biomarkers?” *Current Opinion in HIV and AIDS* 5, no. 6 (November 2010): 463–66. doi:10.1097/COH.0b013e32833ed177.

Suárez, Yajaira, Carlos Fernández-Hernando, Jordan S. Pober, and William C. Sessa. “Dicer Dependent microRNAs Regulate Gene Expression and Functions in Human Endothelial Cells.” *Circulation Research* 100, no. 8 (April 27, 2007): 1164–73. doi:10.1161/01.RES.0000265065.26744.17.

Sugito, Nobuyoshi, Hideyuki Ishiguro, Yoshiyuki Kuwabara, Masahiro Kimura, Akira Mitsui, Hiroki Kurehara, Takuya Ando, et al. “RNASEN Regulates Cell Proliferation and Affects Survival in Esophageal Cancer Patients.” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 12, no. 24 (December 15, 2006): 7322–28. doi:10.1158/1078-0432.CCR-06-0515.

Tabor, S., and C. C. Richardson. “A Single Residue in DNA Polymerases of the Escherichia Coli DNA Polymerase I Family Is Critical for Distinguishing between Deoxy- and Dideoxyribonucleotides.” *Proceedings of the National Academy of Sciences of the United States of America* 92, no. 14 (July 3, 1995): 6339–43.

Tahara, Tomomitsu, Masaaki Okubo, Tomoyuki Shibata, Tomohiko Kawamura, Kazuya Sumi, Takamitsu Ishizuka, Mitsuo Nagasaka, et al. "Association between Common Genetic Variants in Pre-microRNAs and Prognosis of Advanced Gastric Cancer Treated with Chemotherapy." *Anticancer Research* 34, no. 9 (September 2014): 5199–5204.

Tan, B. R., F. Thomas, R. J. Myerson, B. Zehnbauer, K. Trinkaus, R. S. Malyapa, M. G. Mutch, et al. "Thymidylate Synthase Genotype-Directed Neoadjuvant Chemoradiation for Patients With Rectal Adenocarcinoma." *Journal of Clinical Oncology* 29, no. 7 (March 1, 2011): 875–83. doi:10.1200/JCO.2010.32.3212.

Thomson, J. Michael, Martin Newman, Joel S. Parker, Elizabeth M. Morin-Kensicki, Tricia Wright, and Scott M. Hammond. "Extensive Post-Transcriptional Regulation of microRNAs and Its Implications for Cancer." *Genes & Development* 20, no. 16 (August 15, 2006): 2202–7. doi:10.1101/gad.1444406.

Tong, Zan, Xuexian O. Yang, Huichao Yan, Weihuang Liu, Xiaoyin Niu, Yun Shi, Wenfeng Fang, Bing Xiong, Yu Wan, and Chen Dong. "A Protective Role by Interleukin-17F in Colon Tumorigenesis." *PloS One* 7, no. 4 (2012): e34959. doi:10.1371/journal.pone.0034959.

Tosolini, Marie, Amos Kirilovsky, Bernhard Mlecnik, Tessa Fredriksen, Stéphanie Mauger, Gabriela Bindea, Anne Berger, et al. "Clinical Impact of Different Classes of Infiltrating T Cytotoxic and Helper Cells (Th1, th2, Treg, th17) in Patients with Colorectal Cancer." *Cancer Research* 71, no. 4 (February 15, 2011): 1263–71. doi:10.1158/0008-5472.CAN-10-2907.

Trakarnsanga, Atthaphorn, Mithat Gönen, Jinru Shia, Garrett M. Nash, Larissa K. Temple, José G. Guillem, Philip B. Paty, et al. "Comparison of Tumor Regression Grade Systems for Locally Advanced Rectal Cancer after Multimodality Treatment." *Journal of the National Cancer Institute* 106, no. 10 (October 2014). doi:10.1093/jnci/dju248.

Valentini, Vincenzo, Ruud G. P. M. van Stiphout, Guido Lammering, Maria A. Gambacorta, Maria C. Barba, Marek Bebenek, Franck Bonnetain, et al. "Selection of Appropriate End-Points (pCR vs 2yDFS) for Tailoring Treatments with Prediction Models in Locally Advanced Rectal Cancer." *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 114, no. 3 (March 2015): 302–9. doi:10.1016/j.radonc.2015.02.001.

van Meeteren, Laurens A., Marie-José Goumans, and Peter ten Dijke. "TGF-β Receptor Signaling Pathways in Angiogenesis; Emerging Targets for Anti-Angiogenesis Therapy." *Current Pharmaceutical Biotechnology* 12, no. 12 (December 2011): 2108–20.

van Zon, A., M. H. Mossink, R. J. Scheper, P. Sonneveld, and E. a. C. Wiemer. "The Vault Complex." *Cellular and Molecular Life Sciences: CMLS* 60, no. 9 (September 2003): 1828–37. doi:10.1007/s00018-003-3030-y.

van Zon, A., M. H. Mossink, M. Schoester, G. L. Scheffer, R. J. Scheper, P. Sonneveld, and E. A. Wiemer. "Multiple Human Vault RNAs. Expression and Association with the Vault Complex." *The Journal of Biological Chemistry* 276, no. 40 (October 5, 2001): 37715–21. doi:10.1074/jbc.M106055200.

Vasudevan, Shobha, Yingchun Tong, and Joan A. Steitz. "Switching from Repression to Activation: microRNAs Can up-Regulate Translation." *Science (New York, N.Y.)* 318, no. 5858 (December 21, 2007): 1931–34. doi:10.1126/science.1149460.

Vecchio, Fabio Maria, Vincenzo Valentini, Bruce D. Minsky, Gilbert D. A. Padula, Ennapadam S. Venkatraman, Mario Balducci, Francesco Miccichè, et al. "The Relationship of

Pathologic Tumor Regression Grade (TRG) and Outcomes after Preoperative Therapy in Rectal Cancer." *International Journal of Radiation Oncology, Biology, Physics* 62, no. 3 (July 1, 2005): 752–60. doi:10.1016/j.ijrobp.2004.11.017.

Vincenzi, Bruno, Alice Zoccoli, Gaia Schiavon, Michele Iuliani, Francesco Pantano, Emanuela Dell'Aquila, Raffaele Ratta, et al. "Dicer and Drosha Expression and Response to Bevacizumab-Based Therapy in Advanced Colorectal Cancer Patients." *European Journal of Cancer* 49, no. 6 (April 2013): 1501–8. doi:10.1016/j.ejca.2012.11.014.

Wang, Chao-Jie, Johannes Stratmann, Zong-Guang Zhou, and Xiao-Feng Sun. "Suppression of microRNA-31 Increases Sensitivity to 5-FU at an Early Stage, and Affects Cell Migration and Invasion in HCT-116 Colon Cancer Cells." *BMC Cancer* 10 (2010): 616. doi:10.1186/1471-2407-10-616.

Wang, G., Yu, Y., Sun, C., Liu, T., Liang, T., Zhan, L., Lin, X., and Feng, X.H. "STAT3 selectively interacts with Smad3 to antagonize TGF-β."*Oncogene* (November 30, 2015) doi: 10.1038/onc.2015.446.

Wang, Shu-mei, Lu-lu Sun, Wei-xin Zeng, Wan-shui Wu, and Guo-liang Zhang. "Effects of a microRNA Binding Site Polymorphism in SLC19A1 on Methotrexate Concentrations in Chinese Children with Acute Lymphoblastic Leukemia." *Medical Oncology (Northwood, London, England)* 31, no. 7 (July 2014): 62. doi:10.1007/s12032-014-0062-0.

Weinmann, Lasse, Julia Höck, Tomi Ivacevic, Thomas Ohrt, Jörg Mütze, Petra Schwille, Elisabeth Kremmer, Vladimir Benes, Henning Urlaub, and Gunter Meister. "Importin 8 Is a Gene Silencing Factor That Targets Argonaute Proteins to Distinct mRNAs." *Cell* 136, no. 3 (February 6, 2009): 496–507. doi:10.1016/j.cell.2008.12.023.

West, Nathan R., Sarah McCuaig, Fanny Franchini, and Fiona Powrie. "Emerging Cytokine Networks in Colorectal Cancer." *Nature Reviews. Immunology* 15, no. 10 (October 2015): 615–29. doi:10.1038/nri3896.

Wu, Ligang, Jihua Fan, and Joel G. Belasco. "MicroRNAs Direct Rapid Deadenylation of mRNA." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 11 (March 14, 2006): 4034–39. doi:10.1073/pnas.0510928103.

Xu, X. T., Q. Xu, J. L. Tong, M. M. Zhu, F. Nie, X. Chen, S. D. Xiao, and Z. H. Ran. "MicroRNA Expression Profiling Identifies miR-328 Regulates Cancer Stem Cell-like SP Cells in Colorectal Cancer." *British Journal of Cancer* 106, no. 7 (March 27, 2012): 1320–30. doi:10.1038/bjc.2012.88.

Yan, H., Q.-L. Wu, C.-Y. Sun, L.-S. Ai, J. Deng, L. Zhang, L. Chen, et al. "piRNA-823 Contributes to Tumorigenesis by Regulating de Novo DNA Methylation and Angiogenesis in Multiple Myeloma." *Leukemia* 29, no. 1 (January 2015): 196–206. doi:10.1038/leu.2014.135.

You, Kai-yun, Rong Huang, Lu-ning Zhang, Pei-rong Ding, Wei-wei Xiao, Bo Qiu, Hui Chang, Zhi-fan Zeng, Zhi-zhong Pan, and Yuan-hong Gao. "Tailored Selection of the Interval between Neoadjuvant Chemoradiotherapy and Surgery for Locally Advanced Rectal Cancer: Analysis Based on the Pathologic Stage or Chemoradiation Response." *Journal of Cancer Research and Clinical Oncology* 141, no. 4 (April 2015): 719–28. doi:10.1007/s00432-014-1843-8.

Yu, Guopeng, Qianyi Xiao, Xiao-Pin Ma, Xubo Chen, Zhuqing Shi, Lu-Yao Zhang, Haitao Chen, et al. "miR-492G>C Polymorphism (rs2289030) Is Associated with Overall Survival of Hepatocellular Carcinoma Patients." *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, January 11, 2016. doi:10.1007/s13277-015-4752-9.

Zeng, Yan, Li-ke Qu, Lin Meng, Cai-yun Liu, Bin Dong, Xiao-fang Xing, Jian Wu, and Cheng-chao Shou. "HIWI Expression Profile in Cancer Cells and Its Prognostic Value for Patients with Colorectal Cancer." *Chinese Medical Journal* 124, no. 14 (July 2011): 2144–49.

Zeng, Yunhui, Qiongwen Zhang, Hong Wang, Minxun Lu, Hongyu Kong, Yingyi Zhang, and Huashan Shi. "Prognostic Significance of Interleukin-17 in Solid Tumors: A Meta-Analysis." *International Journal of Clinical and Experimental Medicine* 8, no. 7 (2015): 10515–36.

Zhang, Chengyue, Limin Shu, and Ah-Ng Tony Kong. "MicroRNAs: New Players in Cancer Prevention Targeting Nrf2, Oxidative Stress and Inflammatory Pathways." *Current Pharmacology Reports* 1, no. 1 (February 2015): 21–30. doi:10.1007/s40495-014-0013-7.

Zhang, Jun-Xia, Lei Han, Zhao-Shi Bao, Ying-Yi Wang, Lu-Yue Chen, Wei Yan, Shi-Zhu Yu, et al. "HOTAIR, a Cell Cycle-Associated Long Noncoding RNA and a Strong Predictor of Survival, Is Preferentially Expressed in Classical and Mesenchymal Glioma." *Neuro-Oncology* 15, no. 12 (December 2013): 1595–1603. doi:10.1093/neuonc/not131.

Zhang, Lin, Jia Huang, Nuo Yang, Joel Greshock, Molly S. Megraw, Antonis Giannakakis, Shun Liang, et al. "microRNAs Exhibit High Frequency Genomic Alterations in Human Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 24 (June 13, 2006): 9136–41. doi:10.1073/pnas.0508889103.

Zhang, Xinna, Guohui Wan, Franklin G. Berger, Xiaoming He, and Xiongbin Lu. "The ATM Kinase Induces MicroRNA Biogenesis in the DNA Damage Response." *Molecular Cell* 41, no. 4 (February 2011): 371–83. doi:10.1016/j.molcel.2011.01.020.

Zhang, Yingze, Daniel Handley, Tommy Kaplan, Haiying Yu, Abha S. Bais, Thomas Richards, Kusum V. Pandit, et al. "High Throughput Determination of TGFβ1/SMAD3 Targets in A549 Lung Epithelial Cells." Edited by Magnus Rattray. *PLoS ONE* 6, no. 5 (May 20, 2011): e20319. doi:10.1371/journal.pone.0020319.

Zou, Weiping, and Nicholas P. Restifo. "T(H)17 Cells in Tumour Immunity and Immunotherapy." *Nature Reviews. Immunology* 10, no. 4 (April 2010): 248–56. doi:10.1038/nri2742.

Websites

1) *DbSNP Home Page*. NCBI, Web. 18 Mar. 2016.
2) "HaploReg V2." *HaploReg V2*. Web. 18 Mar. 2016.
3) *HapMap Homepage*. Web. 18 Mar. 2016.
4) "MicroCosm Targets Version 5." *Microcosm Targets*. Web. 18 Mar. 2016.
5) "MicroRNA.org - Targets and Expression." *MicroRNA.org*. Web. 18 Mar. 2016.
6) "MicroSNiPer." *MicroSNiPer*. Web. 18 Mar. 2016.
7) "National Institute of Environmental Health Sciences." *SNP Function Prediction*. Web. 18 Mar. 2016.
8) "National Institute of Environmental Health Sciences." *SNPinfo Web Server*. Web. 18 Mar. 2016.
9) "Patrocles Targets Database." Web. 18 Mar. 2016.

10) "The Pharmacogenomics Knowledge Base." *PharmGKB*. Web. 18 Mar. 2016.

11) "PolymiRTS Database." *PolymiRTS Database*. Web. 18 Mar. 2016.

12) "TargetScanHuman 7.0." *TargetScanHuman 7.0*. Web. 18 Mar. 2016.

13) "UCSC Genome Browser Home." *UCSC Genome Browser Home*. Web. 18 Mar. 2016.

14) "Using PubMed." *National Center for Biotechnology Information*. U.S. National Library of Medicine, Web. 18 Mar. 2016.

15) "Welcome To PicTar." *PicTar*. Web. 18 Mar. 2016.