A network biology approach to breast and colorectal cancers

Rodrigo Arroyo Sánchez

TESIS DOCTORAL

UNIVERSITAT DE BARCELONA
FACULTAT DE BIOLOGIA
PROGRAMA DE DOCTORAT EN BIOMEDICINA
2014

A network biology approach to breast and colorectal cancers

Estudio del cáncer de mama y colorrectal mediante biología de redes

Memoria presentada por

Rodrigo Arroyo Sánchez

Para optar al título de doctor por la Universitat de Barcelona realizada en el Institut de Recerca Biomèdica de Barcelona.

DIRECTOR

Dr. Patrick Aloy Calaf

DOCTORANDO

Rodrigo Arroyo Sánchez

A mi familia y a Xey, por estar siempre ahí.

Los imposibles de hoy serán posibles mañana.
Konstantin Tsiolkovsky (1857-1935)

ACKNOWLEDGMENTS

M'agradaria començar aquesta secció donant les gràcies al meu director de tesi, en Patrick Aloy, per donar-me l'oportunitat de realitzar el doctorat i per haver-me format com a estudiant durant aquests anys. Gràcies per haver-me donat tots els medis per fer aquesta tesi de la millor manera i per la teva supervisió.

També vull agrair a la Montse Soler tota la feina feta durant aquests anys com a directora a l'EBL. Gràcies per la teva ajuda, el teu suport i també la teva paciència i comprensió. No era fàcil fer rutllar un laboratori tan peculiar com el nostre, i cal dir que si molts hem arribat fins al final en part es gràcies a tu.

Gracias a los miembros del comité de evaluación de la tesis durante estos años: Travis Stracker, Xavier Salvatella y Josep Lluís Gelpí. También quiero agradecer de antemano a Baldo Oliva, Miquel Àngel Pujana y Carles Ciudad por acceder a evaluar este trabajo, y a los miembros suplentes del tribunal: Timothy Thomson, Modesto Orozco y Ana María Rojas.

Y volviendo al EBL, me faltarían páginas de agradecimientos. Comenzaré por la persona de la que creo haber aprendido más durante estos años; Guillermo. Luego vino su relevo, Nahuai, más de lo mismo, buenos compañeros y mejores personas. Moltes gràcies també a tu Victor per tota la teva ajuda, això si, no puc agarir-te l'estrés que m'han provocat les nostres xerrades futboleres ☺. Obligada mención a parte merecen las dos chicas que me han acompañado de principio a fin en esta experiencia: gracias Chiara y Ozgen por vuestra genial compañía durante todos estos años. Y como no agradecer a los demas miembros pasados o presentes, por el genial ambiente creado entre todos: Clara, Ricart, Elisa, Kathryn, Maica, Marianela, Nuria, Eva, Jofre, Antonio, Montse, Claudio, Cristina y Raquel. Muchas gracias y mucha suerte a ti también Isabelle, estoy seguro que harás genial la faena que tienes por delante.

Obligatorio acordarme también de los compañeros bioinformáticos del grupo. En especial Manu y Miquel, mis compañeros de batallas dentro y fuera del Parc, y también agradecer a Samira su ayuda y apoyo estos últimos y duros meses de tesis. Gracias también a Roland, Amelie, Andreas, Arnaud, Roberto, Lydia, Jacopo, Roger, Laura y Teresa por todos esos buenos almuerzos que hemos pasado juntos.

También quiero acordarme de todos los compañeros de los diferentes equipos de la liga de futbol del IRB, especialmente los más ilustres Nico, Rodo, Evarist, Francesc, Victor B, Radek, Camille, y los ya mencionados Manu, Chiara, Nahuai y Victor.

Quiero también dar las gracias sin duda a todos los grandes amigos que tengo. Gracias a los científicos barceloneses con los que tanto congenio (Alex, Cami, Berni, Roser). Gràcies als empordanesos/esplaiencs/explaiencs per haver acollit a un merengón de la meseta ☺ (Pau, Guti, Alex, Iñaki, Miki, Martí, Jorgina, Joan, Piris, Juli, Guille, Jofre, Genís, Vice, …). Y como no, mencionar a los abulenses, muchas gracias a las familias de Navarre y de Navatal, a mi segunda familia (los R.S.O.), en especial al trío que me acompaña en todas las escapadas (Pablo, Pepe y Marcos). También dar las gracias en general a todos los amigos del pueblo, porque la semana que paso allí es la mejor del año y porque no hay nada más bonito que estar tan sumamente orgulloso de la gente y de la tierra que te vió nacer.

Xey, cariño, a ti muchas gracias por tu alegría, tu compañía, tu apoyo, tus sonrisas, tu ayuda, … en definitiva gracias ser una persona tan especial. Gracias por todos los buenos momentos que hemos pasado juntos y por los que nos quedan por pasar.

Y por último dar las gracias a mi familia porque conforta saber que siempre están ahí. A mi hermana, con quien tantas cosas he compartido. Gracias a mis padres por su sacrificio y por enseñarme lo que son la honestidad y la humildad, y en especial a mi madre, por todo el apoyo, cariño y dedicación que he recibido desde siempre.

Esta tesis va dedicada a todos vosotros.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AD	Activation Domain
BC	Breast Cancer
BC-PIN	Breast Cancer - Protein Interaction Network
BCT	Benefit of Candidate Target
BER	Base Excision Repair
CCLE	Cancer Cell Line Encyclopedia
CI	Crosstalk Inhibition
CK	Cytokeratin
Co-IP	Co-Immunoprecipiation
CRC	Colorrectal Cancer
DBD	DNA-Binding Domain
DC	Drug Combination
DCDB	Drug Combination Database
DCI	Drug Combination Index
DDR	DNA Damage Response
DSB	Double-Strand Break
EM	Expectation-Maximization
EMT	Epithelial to Mesenchymal Transition
FAP	Faminial Adenomatous Polyposis
FDA	Food and Drug Administration
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HC	High-Confidence
HNPCC	Hereditary Nonpolyposis Colorectal Cancer
HRR	Homologous Recombination Repair
IC	Inhibitory Concentration

ICGC	International Cancer Genome Consortium
ICL	Inter-strand Crosslinks
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
IMEx	International Molecular Exchange consortium
IR	Ionizing Radiation
IRIF	IR-induced foci
LOH	Loss of Heterozygosity
MMR	Mismatch Repair
NCI	National Cancer Institute
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
PPI	Protein-Protein Interactions
RND	Random
SD	Standard Deviation
SERM	Selective Estrogen-Receptor Modulator
SSB	Single-Strand Break
TAP	Tandem Affinity Purification
TCGA	The Cancer Genome Atlas
TDC	Target-Drug Combination
TF	Transcription Factor
TTD	Therapeutic Target Database
UAS	Upstream Activation Site
Y2H	Yeast-Two HybridAA fddsfs

GLOBAL INTRODUCTION

In the past century, biomedical sciences have been traditionally immersed in a conceptual reductionism primarily focused on the study of individual molecules. Decades of research into cellular, molecular and structural biology have significantly increased our understanding of the individual proteins taking part in biological processes. However, biological systems are complex in nature, and the knowledge of the components reveals relatively little about their function and organization. As proteins rarely act alone, the traditional approach is unable to predict the behavior of an intact organism and how it coordinately changes in response to a particular stimulus, such as the onset of a disease.
Pharmacological sciences have followed a similar course, with traditional approaches centered on the study, at the molecular level, of the target-compound pair. Many promising drug candidates have resoundingly failed the last clinical phases because the action mechanisms of the pathways they target are still unknown (Pammolli et al., 2011). These effects have been accentuated during the last decade, when the pharmaceutical research has focused on ever more complex diseases that are poorly understood, such as cancer.

Breast cancer (BC) and colorectal cancer (CRC) are a perfect example of a very complex disease that, despite many years of research, is far from being well understood. BC is the most diagnosed cancer in women worldwide and remains the second leading cause of cancer deaths in western countries (Ferlay et al., 2013). Distinct biological features and clinical behaviors turn cancer into a very heterogeneous disease (Weigelt and Reis-Filho, 2009) and, although significant advances in the fight against breast and colorectal cancers have been achieved during the last decades, the current understanding of their biology is still limited.

Studies of gene expression pattern, mutational status, DNA copy number variation, and other protein changes occurring in breast carcinomas have been

performed using both tumor samples and cell lines (Barretina et al., 2012; Ellis et al., 2012; Garnett et al., 2012; Perou et al., 2000). These studies have enormously increased the knowledge of the players involved in breast cancer progression. However, cancer heterogeneity rarely originates from abnormalities in single genes, but rather reflects the disruption of whole complex intra- and intercellular processes (Barabasi et al., 2011). Therefore, the scientific community is rapidly moving to systems approaches, where global properties are considered.

A comprehensive way to describe cancer heterogeneity is the use of network studies. Network biology is a network-based discipline that studies the interactions among molecules, and its focus is on protein-protein interaction (PPI) networks. PPI network-based approaches allow taking proteins back to their context, considering a much broader perspective of their environment without losing the molecular details.

But the availability of the complete map of protein interactions that can occur in a living organism (interactome) is crucial to conduct PPI network analyses. This means that most efforts to unveil the molecular bases of disease pathologies should involve an initial interaction discovery step. For this reason, systematic identification of interactions for a given disease is critical, and high-throughput approaches are being undertaken to enlarge the interactomes (Giot et al., 2003; Hauser et al., 2014; Rajagopala et al., 2014; Rual et al., 2005; Simonis et al., 2009; Stelzl et al., 2005; Uetz et al., 2000). Among many strategies that are up to this task, yeast two-hybrid is one of the most successfully large-scale applied technologies (Fields and Song, 1989).

Among the multiple potential biological and clinical applications of systematic studies based on PPI networks, we are interested in the prediction of protein function (Sharan et al., 2007; Vazquez et al., 2003). There are two types of approaches to infer protein function via a network of interactions: direct annotation schemes, which infer the function of a protein based on its connections within the network, and module-assisted schemes (Sharan et al., 2007). A module-assisted method, as used in this project, first identifies modules of related proteins and, then, annotates each module based on the known functions of its members.

Another area where the analysis of networks can help is in cancer therapy. As research in this field advances, it becomes more and more obvious that the quest for a 'magic bullet' that will defeat all forms of cancers, as antibiotic do against bacterial infections, will never exist. To be effective, novel strategies should avoid the reductionist approach that implicitly assumes that destruction of cancer cells can be achieved by just interfering with a single protein. Different studies and clinical observations have shown that cellular systems are redundant and robust (Kitano, 2004), and cancer cells can find ways to escape a single point blockade. Certainly, treatment failure induced by drug resistance remains a major challenge in most advanced solid cancers, such as breast cancer (Raguz and Yague, 2008).

Taken together, network and systems biology disciplines could revolutionize the study of complex diseases, such as breast and colorectal cancers. If successful, these approaches could significantly help in the development of novel therapeutics for these complex diseases.

This thesis is a multidisciplinary work involving several approaches and relating different research fields, and we have divided the content in two main chapters for a more comprehensive reading. Firstly, we introduce, present and discuss the results obtained regarding the molecular characterization of breast and colorectal cancers interactomes. The following chapter covers the project related to the identification and validation of novel breast cancer drug targets and drug combinations. Finally, we explain the general discussion and conclusions and we present the materials and methods.

CHAPTER

I

MOLECULAR CHARACTERIZATION OF BC AND CRC INTERACTOMES

1. Introduction I

1.1 Global Cancer Burden

Cancer is a term used for diseases characterized by uncontrolled cell growth. Since 1965, the World Health Organization has studied the global burden of cancer. During this time, the pattern of cancer incidence has shifted from occurring predominantly in western nations to becoming a global disease. It was estimated that in 2012 there were 14.1 million new cancer cases diagnosed, 8.2 million cancer-related deaths and 32.6 million people living with cancer (within five years of diagnosis) (Ferlay et al., 2013). This makes invasive cancer the leading cause of death in the developed world and the second leading cause of death in the developing world.

Cancer causes tumors that can expand locally invading nearby parts of the body and may also spread to more distant parts and disseminate systemically. While normal cells are controlled by regulatory signals, cancer cells have the ability to proliferate uncontrolled, invade surrounding tissue and metastasize to distant organs. Cancer commonly starts with mutations in the DNA which can be caused by radiation, chemicals, viruses as well as errors during DNA replication. Hanahan and Weinberg suggested that the complexity of cancer can be reduced to a manifestation of six essential physiologic changes that collectively dictate malignant grow: self-sufficiency in growth signals, insensitivity to antigrowth signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg, 2000). Conceptual progress in the last decade has added two new hallmarks to the list: evading immune destruction, and reprogramming of energy metabolism, in addition to two enabling characteristics: genome instability and tumor promoting inflammation (Hanahan and Weinberg, 2011) (Figure 1).

Cancer itself is not a single disease, but a collection of diseases that arise in various tissues throughout the body. There are more than 100 different types of cancer and most of them are named after the organ or cell type in which they start. Moreover, the severity of a cancer as well as its treatment options varies significantly depending in the tissue of origin. In fact, even within the same cancer type, sub-groups of tumors can be defined which are characterized by unique clinical manifestations (see Section 1.1.1).

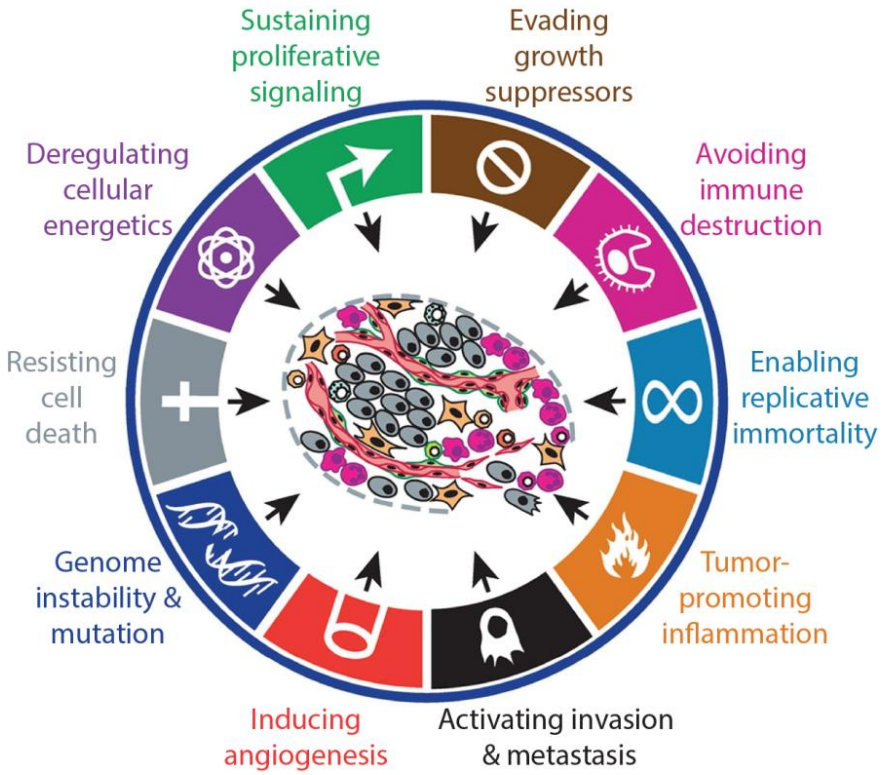


Figure 1. The hallmarks of cancer

Acquired functional capabilities necessary for tumor growth and progression. *Adapted from Hanahan and Weinberg, 2011.*

1.1.1 Breast Cancer

Breast cancer is a type of cancer originating from breast tissue. The most common breast cancers are developed from epithelial cells in the ducts (70-80%) and lobules (10%), known as invasive ductal carcinoma (IDC), and invasive lobular carcinoma (ILC), respectively. Less commonly, breast cancer can begin in the stromal tissues, which include the fatty and fibrous connective tissues of the breast. One of the greatest challenges faced by clinicians and researchers is that breast cancer is not a single entity, but rather a heterogeneous group of several subtypes displaying distinct differences in biological and clinical behavior (Polyak, 2011).

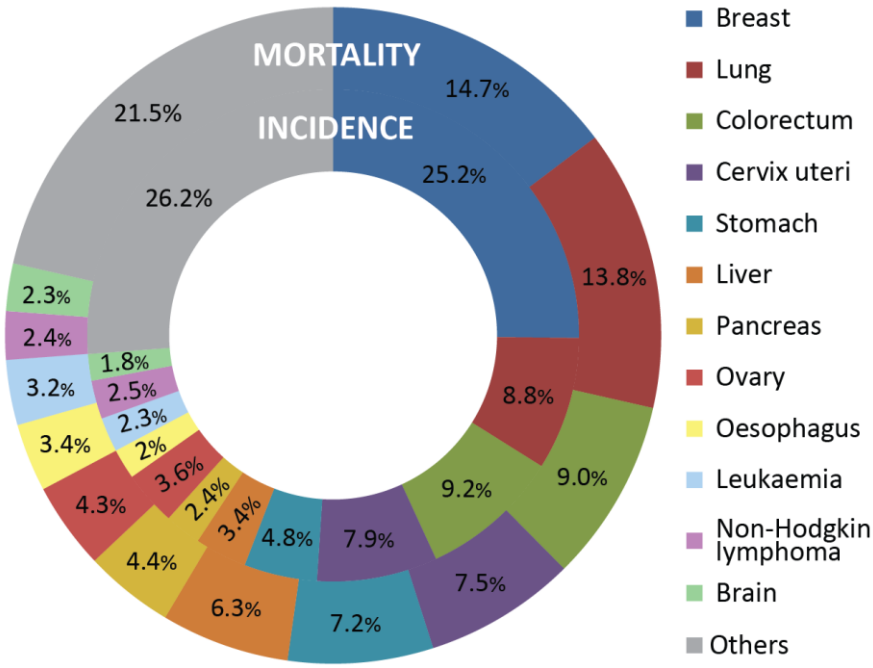


Figure 2. Female cancers incidence and mortality.

Estimated incidence and mortality of female cancers worldwide (2012). Breast cancer is the most frequent cancer among women worldwide (1.67 million new cases diagnosed in 2012, 25% of all cancers) and it is the most frequent cause of cancer death in women worldwide (521,817 deaths, 14.7% of total). *Data source: GLOBOCAN 2012* (Ferlay et al., 2013).

Despite significant improvements in survival over the past 25 years, breast cancer remains as the leading cause of cancer death in women. It is the second most common cancer in the world and, by far, the most frequent cancer among women with an estimated 1.67 million new cancer cases diagnosed in 2012 worldwide (25% of all cancers) (Figure 2).

Initiation and Progression

Breast cancer evolves via sequential progression through defined pathological and clinical stages that are characterized by the acquisition or loss of cellular functions, and altered tissue organization (Polyak, 2007). This process is initiated due to transforming (genetic and epigenetic) events in a single cell within the mammary gland. Subsequent tumor progression is driven by the accumulation of additional genetic changes combined with clonal expansion and selection. Over time, cancer cells can invade healthy breast tissue nearby and spread into the underarm lymph nodes, and then they have a pathway into other parts of the body.

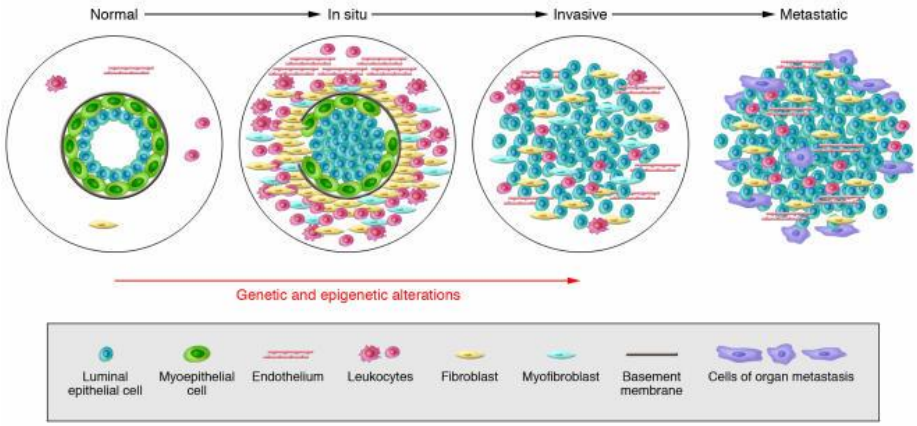


Figure 3. Model of breast tumor progression.

Schematic view cancer progression from normal to metastatic carcinoma progression. *From Polyak, 2007.*

The biology of this tumor progression from in situ to a metastatic stage is illustrated in Figure 3. Normal breast ducts are composed of the basement membrane and a layer of luminal epithelial and myoepithelial cells. In situ

carcinomas are characterized by a multilayered epithelium, due to proliferation of transformed luminal epithelial cells and degradation of the basal membrane. Invasive carcinomas develop due to the loss of myoepithelial cells and the basal membrane, allowing the invasion of tumor cells in surrounding tissues. In situ carcinoma associated myoepithelial cells together with various stromal cells degrade the basement membrane, resulting in early stage invasion of tumor cells. Late-stage invasion and metastatic processes are driven by the accumulation of further genetic and epigenetic aberrations combined with clonal expansion and selection of the most adapted cells (Polyak, 2007).

Apart from the biological changes, breast cancer is commonly separated into different clinical stages which refer to how far the cancer cells have spread beyond the original tumor (Table 1). This staging system provides a strategy for grouping patients with respect to prognosis, as well as therapeutic decisions are formulated in part according to staging categories.

	Stage	Main features
Early stages	Stage 0	Cancer cells remain inside the breast duct, without invasion into normal adjacent breast tissue. This stage is also called lobular carcinoma in situ (LCIS) or ductal carcinoma in situ (DCIS).
	Stage I	Cancer has spread from the lobules or ducts to nearby tissue in the breast. At this stage beyond, breast cancer is considered to be invasive. The tumor is 2 cm or less in diameter and cancer has not spread to the lymph nodes.
	Stage II	The tumor is larger than 2 but no larger than 5 cm. Sometimes lymph nodes may not be involved.
	Stage IIIA	The tumor is 5 cm or larger in diameter, or the tumor may be of any size where cancer cells have spread to axillary lymph nodes.
Advanced stages	Stage IIIB/IIIC	The tumor may be any size but has spread into the skin of the breast, tissues of the chest wall, or lymph nodes near the collarbone. *Inflammatory breast cancer is considered at least stage IIIB.
	Stage IV	The cancer has spread — or metastasized — to other parts of the body, such as bone, liver, lung, or brain.

Table 1. Stages of breast cancer.

Subclassification

The clinical outcome of breast cancer patients is to a large extent driven by the biology of their tumors. There are three receptors; estrogen receptor (ER), progesterone receptor (PR) and epidermal growth factor receptor 2 (ERBB2 or HER2) that have for long guided breast cancer classification. Based on the receptor status, breast cancer subtypes can be defined as i) endocrine receptor (estrogen or progesterone receptor) positive, ii) HER2 positive, iii) triple negative (not positive for estrogen and progesterone receptors and HER2), and iv) triple positive (positive for estrogen receptors, progesterone receptors and HER2).

This classification has been refined by cytokeratin (CK) protein expression patterns. About 70% of primary breast cancers express at least one of the luminal CK proteins (CK7/8/18/19), whereas almost 30% also express at least one of the basal CKs (CK5/6/14). Two minor subtypes express only basal CKs or are negative for both luminal and basal markers, each representing less than 1% of tumors (Abd El-Rehim et al., 2004)

Furthermore, Perou and colleagues used a 4-protein signature that defined four groups of breast cancers: HER2-overexpressing (HER2+), luminal (HER2- and ER+), basal-like (HER2/ER- and CK5+ and/or EGFR+) and the negative group that lacks expression of all four proteins (Nielsen et al., 2004).

In addition to the classifications outlined above, gene expression profiling has defined five major molecular subtypes of breast cancer: luminal A, luminal B, HER2+/ER-, basal-like and normal breast-like (Perou et al., 2000). Luminal tumors are the most common, express ER and have usually a relatively good prognosis. Luminal B cancers differ from the luminal A by having a poorer clinical outcome, being less responsive to the ER antagonist Tamoxifen and having a stronger proliferative signature. The HER2+ subtype does not express hormone receptors and often displays positive lymph nodes at diagnosis. Basal-like tumors have the worst prognosis as they show very low or abolished ER and HER2 expression, while express high proliferation genes (Brenton et al., 2005). Some studies have questioned the existence of the normal-like subtype based on observations that these samples are often associated with low tumor cell percentage (<50%). Consequently, when a tumor sample falls into this

group it does so mostly likely because that sample is predominant composed of normal breast tissue and not tumor tissue. (Prat and Perou, 2011).

The intrinsic subtypes of breast cancer have clinical relevance, as well as the gene expression signatures that had been defined based on clinical outcome of breast cancer patients (van 't Veer et al., 2002) (Figure 4). Although these intrinsic subtypes, for the most part, overlap with immunohistochemically defined subtypes (being ER expression a major classifier, including the luminal A and luminal B subtypes for ER-positive, whereas HER2+, basal-like and normal-like subtypes for ER-negative), they allow for a more precise and clinically meaningful subcategorization of breast tumors.

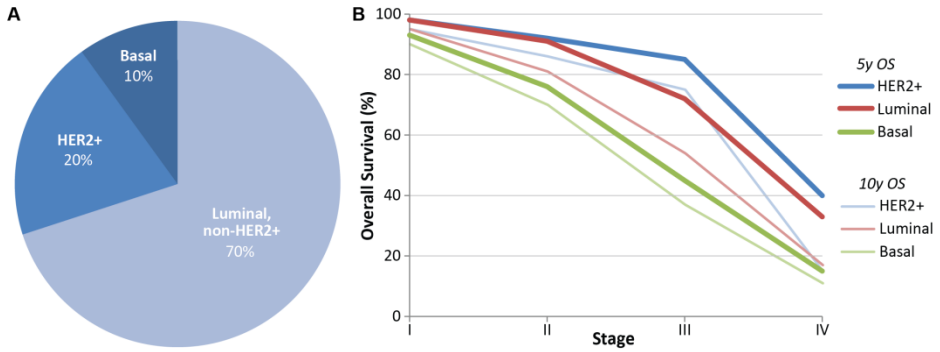


Figure 4. BC subtypes frequency and overall survival.

A) Frequency of breast cancer subtypes. Percentage of Luminal (non-HER2+), HER2+ and Basal-like tumors. **B)** Overall Survival (OS). 5 year and 10 year OS rates of HER2+, Luminal and Basal subtypes at stage I, II, III or IV. *Source: Polyak and Metzger Filho, 2012.*

1.1.2 Colorectal Cancer

Colorectal cancer (CRC) is a cancer from uncontrolled cell growth in the colon or rectum (parts of the large intestine), or in the appendix. Genetic analysis have showns that essentially colon and rectal tumors are genetically the same cancer (Kucherlapati and Wheeler, 2012).

CRC is globally the fourth leading cause of cancer mortality with about 25% of cancers occurring in Europe. In 2012, colorectal cancer accounted for 1.36 million new cases and 694,000 deaths worldwide (Ferlay et al., 2013). The incidence of colorectal cancer is higher in developed countries and, in Spain, it is the most common tumor with approximately 32,240 new cases and 14,799 deaths per year (Ferlay et al., 2013). The disease is rarely diagnosed before an age of 40 and almost non-existent under age 30 (Ferlay et al., 2013).

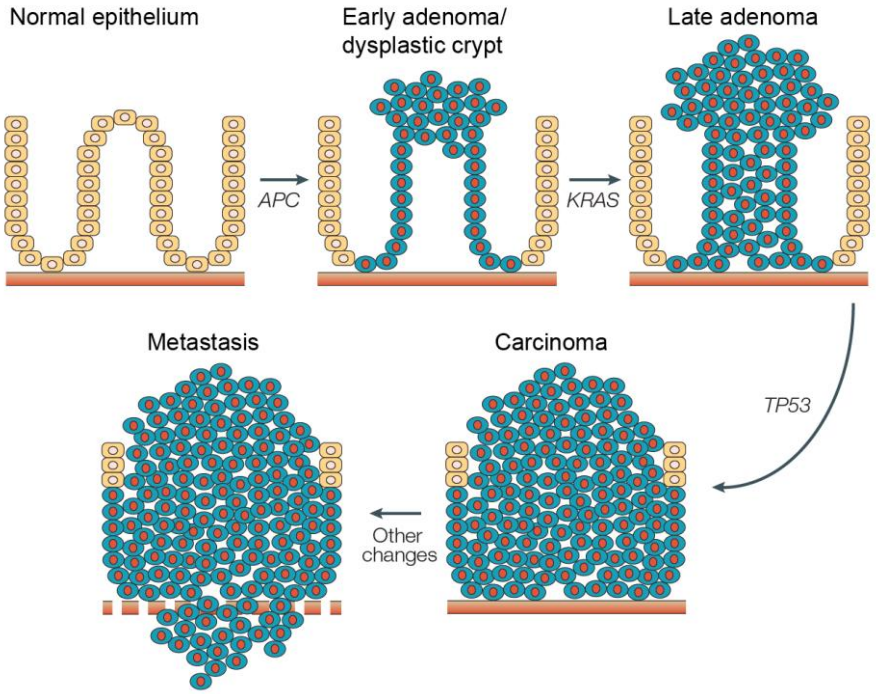


Figure 5. A step-wise model of colorectal tumorigenesis.

This model aligns observed clinicopathological changes with genetic abnormalities in the progression of colorectal cancer (CRC). *From Rajagopalan et al., 2003*.

The normal mucosa of the colon is a highly dynamic system: there is a large amount of proliferating cells which migrate during the differentiation process to the intestinal surface, undergo apoptosis, and are shed into the lumen. In this way, the intestinal epithelium is renewed every five to six days, and this high rate of proliferation in the cells could promote malignant transformation once regulatory mechanisms for cell homeostasis are bypassed. Moreover, the cells of the mucosa are subjected to high toxic and mechanical stress that could affect the cells in their genetic stability. Likely etiologic factors include fecal mutagens, red meat intake, bile acids, altered vitamin and mineral intake and fecal pH (Williams et al., 2011).

Classification of CRC has traditionally been based on histopathological features, and molecular studies have allowed a significant appreciation of the heterogeneous nature of CRC (Kucherlapati and Wheeler, 2012). The essential etiologic element of colorectal cancer can be observed in genetic changes of the epithelial cells in the colonic mucosa. Vogelstein et al. correlated the morphologic changes from normal colonic mucosa and adenomatous polyps to cancer with the accumulation of genetic aberrations (Fearon and Vogelstein, 1990). Whereby, subsequent mutations or inactivation of APC, KRAS, and p53 characterize the progression from aberrant crypts to early and late adenomas and to carcinomas (Rajagopalan et al., 2003; Walther et al., 2009) (Figure 5).

Furthermore, there has been a significant advance in identifying the specific driver genes and pathways important in the initiation and progression of CRC, as well as the constellation of somatic alterations that are present in sporadic CRCs, which include the WNT, RAS-MAPK, PI3K, TGF-b, P53 and DNA mismatch-repair pathways (Fearon, 2011). Large-scale sequencing and expression analyses have identified numerous genes recurrently mutated or whose expression is dysregulated in colorectal tumors (Sjoblom et al., 2006; Wood et al., 2007). However, the role of these altered genes in CRC disease is not clear. Further mechanistic insights into these relationships may enable a deeper understanding of the pathophysiology of CRC and may advance the identification of novel therapeutic targets.

1.1.3 Advances and Challenges of Cancer Research

The scientific community is making a huge effort to find a cure for cancer and to eradicate it as a major cause of death (Ferlay et al., 2013). Despite significant progress in the treatment of certain forms cancer types (Ellis et al., 2012), it remains as a global disease. For this reason, cancer research has been intensified with the aim to improve the understanding of its biology and the development of more effective cancer treatments.

All cancers arise as a result of somatically acquired DNA mutations of tumor cells. That does not mean, however, that all the somatic abnormalities present in a cancer genome have been involved in development of the cancer (Stratton et al., 2009). A driver mutation is causally involved in oncogenesis, and it has provided growth advantage on the cancer cell. In turn, a passenger mutation does not have functional consequences; therefore it has not conferred clonal growth advantage and has not contributed to cancer development. An important challenge of cancer research consists of efficiently recognizing the guilty drivers present in the large datasets of genes altered in cancer.

Major advances in cancer research resulted from the novel next-generation sequencing technologies (Bentley et al., 2008; Campbell et al., 2008). Over the last decade, massively parallel sequencing of cancer genomes has largely been performed, using either frozen tissue samples or immortalized cancer cell lines (Greenman et al., 2007; Wood et al., 2007). This means that a large fraction of genes altered in cancer could be identified at base-pair resolution. A leading force in this endeavor is The Cancer Genome Atlas (TCGA) research network, which aims to explore the entire spectrum of genomic variations involved in human cancer (http://cancergenome.nih.gov/). Similar to TCGA, the International Cancer Genome Consortium (ICGC), another initiative of research projects, attempts to generate comprehensive catalogues of genomic abnormalities in tumors from 50 cancer subtypes (https://www.icgc.org/).

Another methodological advance, DNA microarray technology, initiated an explosion of gene expression analyses (transcriptomics). More than 100 published studies have analyzed gene expression signatures for most major cancer types and subtypes. This has led to the identification of specific gene expression patterns that correlate with various characteristics of tumors

including tumor grade or differentiation state, metastatic potential, and patient survival (Perou et al., 2000; Rhodes et al., 2004).

Apart from DNA/mRNA expression profiles, protein expression profiles have been also used to shed light on changes on the proteomic level. The Human Protein Atlas project, for example, performs an antibody-based approach to generate protein expression profiles for a large number of human tissues, cancers and cell lines (http://www.proteinatlas.org/) (Uhlen et al., 2005). This method allows a comparison of protein expression data of normal and disease tissues.

When focusing on the development of novel cancer drugs, recent advances in RNA interference (RNAi) technologies have made it possible to systematically search for genes whose loss of function yields cell lethality. These advances have been crucial to carry out high-throughput screening in cancer cells, providing novel targets for the next generation of anticancer agents (Luo et al., 2009; Neumann et al., 2006).

In summary, most current follow-up initiatives are directed to study the complexity of cancer, however the identification of the genomic, transcriptomic and proteomic changes associated with each cancer type often does not lead to clinically prosecutable therapeutic strategies. Thus, the knowledge of all the individual genes somehow altered in a given cancer type seems to be not enough. Cancer is a complex disease and, to unravel its underlying mechanisms, systems biology has to be implemented in addition to single gene studies. Systems biology paradigm permits study cancer as a network of associated disease genes (see Section 1.3). Hence, the construction and utilization of genome- and proteome-wide interaction networks can improve the success in cancer research. A pivotal role in this new system biological strategy is the study of protein-protein interaction networks, as protein interactions have long been known to be crucial in oncogenesis (Arkin, 2005). With the generation of comprehensive interaction maps, where also global properties are considered, we can advance in the study and prediction of cancer-associated processes. Although considerable challenges are still to overcome, interactomics evolve as a cornerstone in the systems biology of cancer.

1.2 Genomic Instability

Genome or genomic instability refers to a high frequency of mutations or alterations in the genome of a cellular lineage. These mutations can include changes in nucleic acid sequences, chromosomal rearrangements or aneuploidy. As we discuss in Section 1.2.2, genomic instability is a major driving force for tumorigenesis (Shen, 2012).

One source of genomic instability is endogenous (metabolically-caused), as cellular metabolism generates many chemical products that can react with the DNA molecules altering their structure and function. Moreover, DNA is exposed to exogenous threats ranging from radiation to genotoxic chemical species, whose damage adds up to the endogenous. Since damaged DNA is normally prevented from being replicated and transcribed, a number of mechanisms have evolved to repair the lesions while preventing the cell from progressing through the cell cycle. Therefore, another source of genome instability may be epigenetic or mutational reductions in expression of DNA repair genes.

Most of these repair processes are carried out in a multi-step fashion by enzymatic complexes encompassing elements that i) recognize the damage and recruit the repair proteins, ii) reverse the lesion and iii) prevent the cell to progress in the cell cycle with unrepaired damages. Some repair mechanisms are very lesion-specific, while others are partially redundant and can resolve a broader range of damage types.

1.2.1 DNA Repair Mechanisms

In general DNA repair can be divided into pathways that repair damage of one of the DNA strands (mismatches, subtle base modifications, bulky adducts, single-stranded breaks or gaps) or damage that affects both DNA strands (crosslinks, double-stranded breaks).

Repair of Single-Strand Breaks

Several DNA repair pathways exist for repair of different types of single-strand breaks (SSBs) such as DNA adducts and mismatched bases. These pathways use the intact complementary DNA strand for error-free repair.

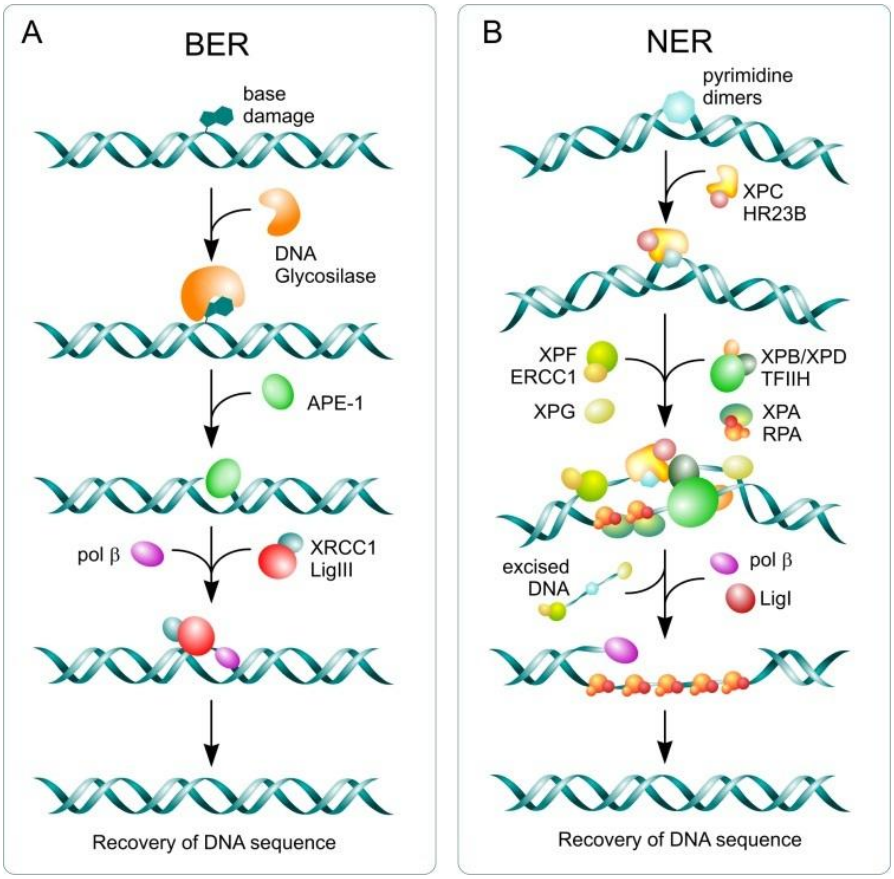


Figure 6. Major pathways in SSBs repair.

A) Schematic representation of the basic steps followed during base excision repair (BER). **B)** Main sequence of events and enzymatic activities implicated in nucleotide excision repair (NER). *From Mladenov and Iliakis, 2011.*

The base excision repair (BER) pathway takes care of non-bulky base modifications in the DNA such as oxidative modifications, methylation, or alkylation. Furthermore, BER repairs SSBs in the DNA, caused by ionizing radiation or by platinum agents. In brief, DNA glycosylases specific for different types of DNA damage cleave the DNA around the damaged base and remove the damaged base from the helix but not from the sugar phosphate backbone. Subsequently, polymerase β fills the single nucleotide gap using the complementary DNA strand as a template and the nick is sealed by the XRCC1-ligase3 complex (Parsons and Dianov, 2013) (Figure 6).

In contrast to BER, nucleotide excision repair (NER) pathway is responsible for clearing helix-distorting lesions from the DNA, such as those induced by ultraviolet radiation or chemotherapeutics causing bulky intra-strand DNA adducts. Using a broad range of proteins, the NER pathway i) unwinds ~30 base pairs of DNA around the damage site through helicases, ii) cleaves the DNA using endonucleases, and iii) fills the resulting gap using the complementary DNA strand as a template (Figure 6) (Cleaver et al., 2009).

The DNA mismatch repair (MMR) is a system that specifically recognizes and repairs erroneous mis-incorporated bases and insertion/deletion loops that can occur during DNA replication. In short, hMSH proteins form heterodimers which recognize the mismatched bases and recruit a protein complex that facilitate excision of the mismatched DNA by an exonuclease. Then polymerase δ and PCNA resynthesize the DNA and the remaining nick in the DNA is sealed by ligase 1 (Jiricny, 2006).

Repair of DNA Double-Strand Breaks

Double-strand breaks (DSBs) are mostly induced by free radicals, ionizing radiation, chemotherapeutics forming DNA inter-strand crosslinks (ICLs) and the conversion of SSBs into DSBs by replication fork collapse during DNA replication (Hoeijmakers, 2001). Unlike in SSBs, in the presence of a DNA double-strand break (DSB), repair systems no longer can depend on the complementary strand for correct repair.

The presence of a DSB is sensed by the MRN complex of MRE11/RAD50/NBS, which localizes to both DNA ends and subsequently recruits ATM, which is responsible for checkpoint activation and cell cycle

arrest through TP53. ATM also phosphorylates histone H2AX (γH2AX) resulting in chromatin remodeling around the break and recruitment of DNA damage response factors such as BRCA1 (Hartlerode and Scully, 2009). Depending on the phase of the cell cycle, DSBs are repaired either by NHEJ, which takes place in G0-G1 phase, or by HRR, which takes place in the S or G2 phase (Figure 7).

Non-homologous end joining (NHEJ) is an error-prone mechanism for ligation of DNA DSBs. In brief, after phosphorylation of γH2AX, a heterodimer of KU70/KU80 binds to both DNA ends and recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). The DNA-PKcs proteins interact on either end of the DSB, forming a bridge between both DNA ends (Hartlerode and Scully, 2009). The MRN complex has been suggested to play an additional role in NHEJ, probably in stabilizing the two DNA ends (Dinkelmann et al., 2009). Lastly, the break needs to be sealed by ligating the DNA ends back together; the complex of XRCC4/Ligase 4 is responsible for this step (Figure 7). Indeed, since NHEJ fuses DNA ends without taking into account the missing DNA or a template, this pathway is error-prone.

In contrast to NHEJ, DNA DSB repair by homologous recombination is error-free, since the homology of the sister chromatid is used for repair. To search for this homology, a long 3'end DNA overhang needs to be created. For this, the MRN complex is needed again, which interacts with CtIP, EXO1 and the BLM helicase (Hartlerode and Scully, 2009). The created single-stranded DNA ends are subsequently coated with RPA; however, to start the search for sequence homology, RPA needs to be replaced by RAD51. This process is directly mediated by BRCA2 and to facilitate this replacement, a complex of BRCA1/BARD1 needs to be present. The exact interaction remains unknown, but it is thought that PALB2 may connect BRCA2 and BRCA1/BARD1. RAD51 subsequently invades the sister chromatid, resulting in partial displacement of the non-complementary strand (D-loop) (Sy et al., 2009). After alignment of the invading DNA strand with the homologous DNA duplex, the chromatin remodeling functions of RAD54 operate to facilitate DNA synthesis and branch migration. It results in formation of Holliday junctions, enabling DNA synthesis using the sister chromatid as a template. Lastly, the DNA

structures formed by the D-loop or Holliday junction are resolved by resolvases such as BLM, and GEN1 (Hartlerode and Scully, 2009) (Figure 7).

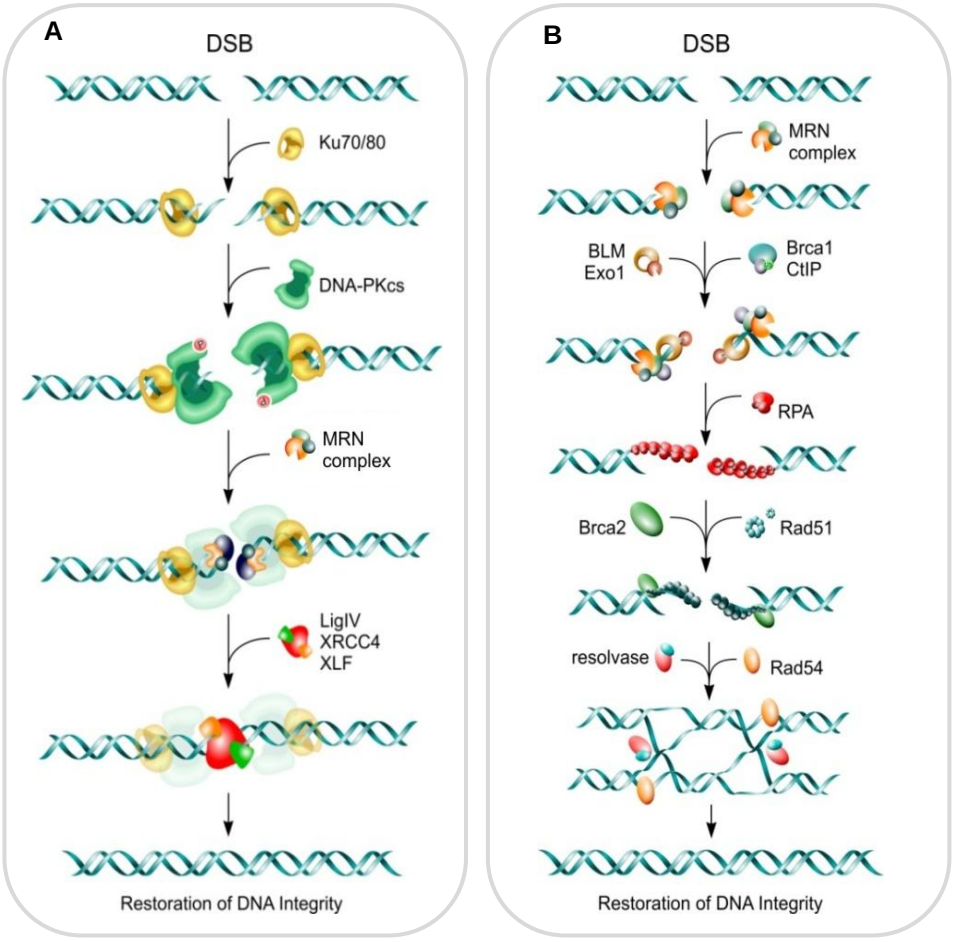


Figure 7. Major pathways in DSBs repair.

A) Repair of DSBs by non-homologous end joining (NHEJ). Major enzymatic activities involved in the repair of DSBs by the simple joining of the free DNA ends are depicted. **B) Homologous recombination repair (HRR) of DSBs.** Main steps involved in HRR are illustrated. *From Mladenov and Iliakis, 2011.*

1.2.2 Genomic Instability and Cancer

Genome instability is central to carcinogenesis. The link between genomic instability and cancer is well established, since many inherited defects in DNA repair genes lead to genomic instability and predispose to malignancies, such as breast cancer. This illustrates the importance of DNA repair pathways for maintaining genomic integrity and preventing cancer (Hoeijmakers, 2001).

Given the redundancy in pathways that prevent overreplication or induce DNA repair, it is unlikely that affecting single pathways create so much genomic instability to cause cell death. Instead, inactivation of these single pathways may simply increase the mutation burden without provoking cell death. The resulting heterogeneity in the genes of daughter cells allows the appearance of cells with growth or survival advantages, i.e., the driving force for cancer development.

A similar genetic heterogeneity underlies the resistance mechanisms of cancers to many types of therapy. In fact, in many cancers the loss of function in one DNA repair pathway is often partially compensated by the remaining ones. These repair mechanisms are occasionally up-regulated providing the cancer cells with means to better resist genotoxic damages, including those received for therapeutic purposes.

Thus, a major objective should be to exploit known anomalies of repair and replication control in cancers to determine which redundant repair pathways should be therapeutically inhibited to produce 'synthetic lethality' in the cancer cells (Bouwman and Jonkers, 2012). This will convert the mutation-generating machinery in the cancer into an 'Achilles heel' by pushing the malignant cells selectively into extensive genomic instability and cell death (Abbas et al., 2013).

1.2.3 Hereditary Breast Cancer and Genomic Instability

The risk of breast cancer is determined by both genetic and lifestyle factors. Lifestyle factors are mainly related to events affecting hormonal status (for example parity and breast-feeding history, age of menarche and of menopause, use of oral contraceptives) and to environmental agents (ionizing radiation, fat-rich diet, alcohol consumption and so on) (Dumitrescu and Cotarla, 2005). While the variation in breast cancer incidence among populations can be largely explained in terms of these exogenous factors, there is a substantial variation between individuals that is genetically determined.

Approximately 5-10% of the breast cancers are caused by mutations in cancer predisposition genes, such as BRCA1 and BRCA2 (Turnbull and Rahman, 2008). Mutations in these two genes increase the relative risk of breast cancer by 10- to 20-fold, and account for approximately 80-90% of the familial breast cancer cases (Chen and Parmigiani, 2007). They are tumor suppressor genes that, when mutated, lead to the inability to regulate cell death and hence, to uncontrolled cell growth, resulting in cancer. BRCA1 derived breast cancers are more frequently triple negative (basal type), while patients with BRCA2 mutations often develop tumors that are ER and PR positive (luminal phenotype).

Mutations in other breast cancer predisposition genes are known for years, but their impact for breast cancer development remains to be elucidated. These include high penetrance but low frequency mutations, such as TP53 (Li-Fraumeni), PTEN (Cowden), and STK11 (Peutz-Jeghers). Other genetic mutations have recently been identified and are characterized as high frequency but less penetrance mutations, such as CHEK2, ATM, BRIP1, and PALB2 mutations (Byrnes et al., 2008). Specific genetic alterations that have been identified for many of the established hereditary breast cancer syndromes are shown in Table 2 (Thompson and Easton, 2004; Turnbull and Rahman, 2008).

Most of these genes encode proteins involved in DNA damage response, directly linking cancer development to deficiencies in maintaining the integrity of the genetic information. Loss or malfunctioning of genome caretakers leads to an increased probability of acquiring new mutations and, therefore, to gain even more of the hallmarks of cancer previously mentioned in Section 1.1 and

illustrated in Figure 1. Furthermore, genomic aberrations have also been showed to correlate with clinical features and therefore proposed as a possible prognostic marker (Chin et al., 2006).

Penetrance	Gene / Locus	Relative Risk of Breast Cancer	Population Carrier Frequency
High penetrance	*BRCA1*	>10	1 in 860
	BRCA2	>10	1 in 740
	TP53	>10	1 in 5,000
Uncertain penetrance	*PTEN*	2-10	1 in 250,000
	STK11	2-10	1 in 25,000-280,000
	CDH1	2-10	Rare
Intermediate penetrance	*ATM*	2-3	1 in 125
	CHEK2	2-3	1 in 90
	BRIP1	2-3	1 in 250
	PALB2	2-4	1 in 500
Low penetrance	10q26, 16q12, 2q35, 8q24, 5p12	1.08-1.26	24-50%
	11p15, 5q11	1.07-1.13	28-30%
	2q33	1.13	0.87%

Table 2. Breast cancer predisposition factors.

Given the exceptional relevance of DNA repair in breast cancer, we propose that the discovery of novel proteins related to this cellular process should be a priority. This novel DNA repair genes could have powerful applications either in the study of hereditary BC or in the discovery of novel clinically relevant drug targets.

1.3 Systems and Network Biology

Systems biology is an emerging approach that analyzes the relationships among the elements of a system with the objective of understanding its properties. A system may include for example just a few protein molecules, a more complex molecular machine or a cell or group of cells executing a particular function. Thus, systems-wide analysis can be applied to all biological levels; molecules, cells, organs, organisms or even populations and ecosystems. In each case, it is necessary to describe all elements of a system, define the relationships among these elements and characterize the flow of information that links these elements to an emergent biological process (Kitano, 2002).

Systems analysis can be applied to cells. The cell is functioning thanks to complex interactions of several types of biomolecules such as DNA, RNA and poteins. Albeit the function of a single gene might present a molecular description of a cellular phenotype, it is often not sufficient to explain the particular processes. In consequence, many fundamental biological questions remain unanswered as traditional approaches cannot capture the full repertoire of biochemical activities within cells (Jaeger and Aloy, 2012). However, recent advances in biological data collection and bioinformatic techniques had promoted that shift from gene-centric approaches for phenotypic characterization to more systematic approaches (Guney et al., 2012).

Thanks to the integration of various data sources, we can better understand how organisms function with the cooperation of groups of biomolecules that constitute biological systems such as genetic (genomics), epigenetic (epigenomics), metabolic (metabolomics) and PPI (interactomics) systems. All these complex biological systems may be represented and analyzed as computable networks and, consequently, the past decade witnessed a brand new perspective named network biology.

Network biology is the study of these biological networks, which capture a variety of molecular interactions and thus provide an excellent opportunity to consider physiological characteristics of individual molecules within their cellular context (Barabasi and Oltvai, 2004). As diseases derive from alterations in the cellular processes, network biology plays a key role in unveiling disease mechanisms. Indeed, network-based approaches have been used to analyze

biological systems such as gene regulatory networks (Davidson et al., 2002), metabolic networks (Jeong et al., 2000), signal transduction networks (Sambrano et al., 2002), gene co-expression networks (Stuart et al., 2003) and protein interaction networks (Jeong et al., 2001).

1.3.1 Protein-Protein Interactions

Decades of research have produced a remarkable amount of knowledge about the function and molecular properties of individual proteins. In fact, proteins are vital macromolecules, at both cellular and systemic levels, but they rarely act alone. They interact with each other in a highly specific manner, thus protein-protein interactions (PPIs) play a key role in many cellular processes.

The first step needed is to define correctly what PPIs are. Commonly they are understood as physical contacts between two or more proteins that occur in a cell or in a living organism in vivo. But the issue of whether two proteins share a "functional contact" differs from the question of whether the same two proteins interact directly with each other. For example, any protein in the ribosome or in the basal transcriptional apparatus shares a functional contact with the other proteins in the complex, but not all the proteins in the particular complex interact. Therefore, the other types of functional links between biomolecular entities (genes, proteins, metabolites, etc.) should not be confused with physical protein interactions. The physical contacts considered as PPIs should be specific, excluding interactions that a protein experiences when it is being made, folded, or degraded. Therefore, the definition of PPI has to consider (De Las Rivas and Fontanillo, 2010): (1) the interaction interface should be the result of specific selected biomolecular events/forces, and (2) the interaction interface should be non-generic (i.e., evolved for a specific purpose distinct from totally generic functions). It is also important to define the biological context of PPIs. Not all possible interactions will occur in any cell at any time. Instead, interactions depend on cell type, cell cycle, environmental conditions, protein modifications, presence of cofactors, and presence of other binding partners.

As it was said before, PPIs are involved in most cellular processes. Therefore, identifying and characterizing PPIs and their networks is essential for understanding the mechanisms of biological processes on a molecular level.

The complete map of protein interactions that can occur in a living organism is called the interactome. Interactome mapping has become one of the main scopes of current biological research, similar to the way "genome" projects were a driving force of molecular biology 20 years ago (De Las Rivas and Fontanillo, 2010).

Different experimental techniques have been developed to detect PPIs in order to unravel the global picture of protein interactions in a cell (Shoemaker and Panchenko, 2007). For instance, PPIs have been identified for a long time using low-throughput biophysical methods such as NMR or crystallography. However, more recently, much effort was directed towards the development of high-throughput interaction detection methods, which provided sufficiently automatized possibilities to unveil PPIs on a proteome-wide scale. The two most prominent high-throughput techniques include tandem affinity purification (TAP) tagging (Rigaut et al., 1999) and yeast two-hybrid screen (Y2H) (Fields and Song, 1989), with the latter one being used in our project.

1.3.2 The Yeast Two-Hybrid Approach

The yeast two-hybrid screen (Y2H) is a simple and rapid method for finding the interactipartners of a target protein; it identifies proteins that interact as a binary complex with the target protein. Y2H is based on the fact that many eukaryotic transcription factors (TF) have at least two distinct domains, a DNA-binding domain (DBD) that directs binding to a promoter DNA sequence and an activation domain (AD) that activates transcription. Protein-protein interactions can be identified by splitting this protein into two parts; the DBD is fused to the bait protein, whereas the AD is attached to the prey protein. When bait and prey proteins interact in the cell, the DBD and AD of the TF are brought together, resulting in an in vivo reconstitution of the TF, which then can activate the transcription of specific reporter genes (Figure 8) (Fields and Song, 1989).

The screen can use several reporter genes, whose activation is easily detected, such as URA3, HIS3 or lacZ. URA3, for example, encodes orotidine-5'-phosphate decarboxylase, an enzyme required for the biosynthesis of uracil. Therefore, activation of URA3 reporter gene is observed only when the two dimerization domains interact and position the AD in the correct site upstream

of the reporter gene. This URA3 activation allows cells to grow in the selective media lacking uracil (URA3+ phenotype).

The Y2H method has been widely used to uncover a large number of interactions (Koegl and Uetz, 2008). The advantage of Y2H, compared to other techniques, is that it can be used to detect transient PPIs in vivo and in large-scale fashion. However, it has been also widely criticized, mainly because it tends to identify a high proportion of false positives: protein hits that are unlikely to interact with each other in vivo. Among possible reasons for false positive interactions in yeast may be a high expression level of bait and prey and their localization in a compartment which does not correspond to their natural cellular environment, or proteins that are known to be 'sticky' or that are not correctly folded can show unspecific interactions (Bruckner et al., 2009). On the other hand, false negatives in Y2H refer to PPIs which cannot be detected due to limitations of the screening method. For instance, fusion of the AD or DBD domains to the target proteins can alter their interaction interface, resulting in false negative interactions (Aloy and Russell, 2002).

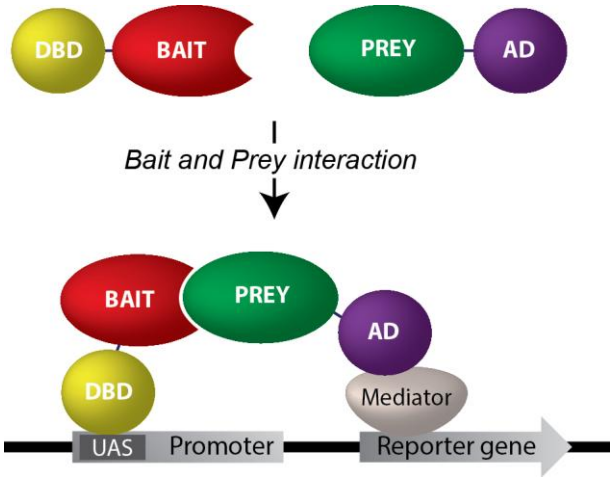


Figure 8. The Yeast Two-Hybrid screen.

A DNA-binding domain that binds to an Upstream Activation Site (UAS) is fused to a bait protein and the prey protein is fused to a transcriptional activation domain. In this way, interaction of bait to prey leads to transcription of the reporter gene.

It could be also good to say here, that the Y2H technique and results have been improved and are not so prone to false-positive anymore compared to its beginnings.

Two Y2H screening approaches can be distinguished: the matrix, which is pairwise array-based, and the library approach, where one can search for pairwise interactions between a defined protein of interest (bait) and their interaction partners (preys) present in cDNA libraries. Hence, as preys are not separated on an array but pooled in libraries, it is largely covering a transcriptome and reducing the rate of false negatives. However, the rate of false positives is increased. In addition, interaction partners have to be identified by colony PCR analysis and sequencing, making library screens more expensive and time consuming.

Despite these limitations, the Y2H system have been improved (Koegl and Uetz, 2008) and it remains a powerful tool for large-scale screening in interactomics. Therefore, Y2H has been applied to several sets of proteins from different organisms, including E. Coli (Rajagopala et al., 2009), yeast (Uetz et al., 2000), worm (Simonis et al., 2009), fly (Giot et al., 2003) and human (Rual et al., 2005; Stelzl et al., 2005).

1.3.3 Protein-Protein Interaction (PPI) Networks

The protein interaction networks allow complex and dynamic responses to diverse cellular events and environmental stimuli. The collective understanding of the structure and nature of protein interaction networks is amongst the best appreciated of biological networks. In the last 15 years, large-scale interaction screens using different experimental techniques have reported thousands of new interactions (Hauser et al., 2014; Rajagopala et al., 2014; Rual et al., 2005; Simonis et al., 2009; Stelzl et al., 2005) which have been deposited in various databases such as BIND (Isserlin et al., 2011), BioGRID (Chatr-Aryamontri et al., 2012), DIP (Salwinski et al., 2004), HPRD (Keshava Prasad et al., 2009), IntAct (Kerrien et al., 2012), MINT (Licata et al., 2012), MIPS (Mewes et al., 2011). The International Molecular Exchange (IMEx) consortium (Orchard et al., 2012) is an international collaboration to share literature-curation efforts and to provide nonredundant sets of protein interactions within a single search interface (http://www.imexconsortium.org/).

Protein interaction networks can be represented as a graph, where the nodes are proteins and the edges represent physical interactions between proteins. The distance between two proteins of the network is defined as the minimum number of edges that one has to follow in order to connect the two proteins. The PPI network for a given protein can be built at different depths, which represents the number of interacting steps that can be taken from the source protein to the outermost protein of the network. Furthermore, in a protein interaction network, we refer to proteins with high connectivity (i.e. with many interaction partners) as 'hubs'. In conclusion, representing protein interactions in a network has fundamental advantages over the traditional approach of storing interaction data in the form of simple lists (Russell and Aloy, 2008).

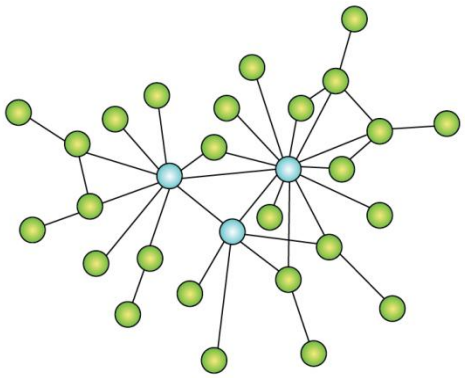


Figure 9. The model of a scale-free network.
Representation of a scale-free network, a network whose degree distribution follows a power law. Hubs are depicted as blue.

Interestingly, interactome networks appear to have a power law degree distribution: most proteins interact with few partners, whereas a few proteins ('hubs') interact with many partners (Yook et al., 2004). Networks following a power law degree distribution are called 'scale-free', due to the absence of a typical node in the network (i.e. no node has a degree that is characteristic for all the other nodes in the network) (Figure 9). Other important global properties observed in PPI networks include the 'small-world effect' (i.e. any two nodes are connected with a path of a few edges only) and a high level of 'modularity' (i.e. interacting proteins tend to be in complexes or act in the same

functional process) (Barabasi and Oltvai, 2004; Spirin and Mirny, 2003). Though groups of genes in a cell are organized to minimize the effects of perturbations, biological systems are prone to be disrupted by certain types of rare but specialized perturbations leveraging the fragility of the system, for example, seizing hub proteins in PPI networks can predominantly break up the network (Albert et al., 2000).

Application of PPI Networks

Apart from being a useful representation of interaction data, systematic studies of PPI networks can be applied to objectives such as: the prediction of protein function (Sharan et al., 2007; Vazquez et al., 2003), the identification of functionally coherent modules (Dittrich et al., 2008), the estimation of interactions reliability (Bader et al., 2004), the identification of domain-domain interactions (Guimaraes et al., 2006), the prediction of protein interactions (Lehner and Fraser, 2004), the study of the relationships between network structure and function (Yook et al., 2004), the detection of proteins involved in disease pathways (Rhodes et al., 2005), the study of pharmacological drug-target relationships (Pujol et al., 2010), and the comparison between model organisms and humans (Gandhi et al., 2006).

Aside from the evident relevance of network analysis at different levels of study, we focused on two particular applications: the identification of functional modules sharing common cellular function, and the outcome of PPI network studies to predict the function of proteins. Network-based approaches for elucidating protein function can be classified in direct methods, which propagate functional information through the network, and module-assisted networks, which infer functional modules within the network and use those for the annotation task (Sharan et al., 2007). The common principle of all direct methods for functional annotation is that proteins that lie closer to one another in the PPI network are more likely to have similar function. On the other hand, module-assisted methods try to identify coherent groups of genes (in terms of network topology or data from experiments) and then assign functions to all the genes in each group.

1.3.4 PPI Networks and Disease

Nowadays, it is evident that complex disease phenotypes rarely originate from abnormalities in single genes but rather reflect the disruption of the complex intra- and intercellular processes that interact in a network (Barabasi et al., 2011). In particular, the full molecular complexity of common human diseases, such as cardiovascular diseases, diabetes, cancer, or neurological disorder, do not obey the standard Mendelian patterns of inheritance, and can only be anticipated by the study of the processes that interact in a complex network. This interconnectivity implies that the impact of a specific genetic abnormality can spread along the links of the network and alter the activity of gene products that otherwise carry no defects. Therefore, an understanding of a gene's network context is essential in determining the phenotypic impact, and elucidating the underlying mechanisms is crucial for understanding the onset of diseases and for the development of disease-specific diagnostic and therapeutic approaches (Jaeger and Aloy, 2012).

Disease network properties have been studied from a global perspective, highlighting a strong association between protein network connectivity and disease. For instance, recent studies demonstrated that proteins encoded by disease genes tend to interact with each other compared to the rest of the proteins (Gandhi et al., 2006) and disease genes whose mutations are somatic (i.e. most cancer types) are more likely to encode for protein hubs (Goh et al., 2007). Based on these findings, a series of network-based tools have been developed to predict potential disease genes (similar to function prediction). These tools exploit the guilt-by-association principle assuming that direct interactors of a disease protein are likely to be involved in similar disease phenotypes (Oti et al., 2006). On the other hand, the module-based methods associate genes to a disease using disease modules and the topology of the PPI network (Navlakha and Kingsford, 2010). In conclusion methods that use global topology of the network have been demonstrated to outperform methods that only take the local or no topology information into account (Navlakha and Kingsford, 2010), and network-based approaches to human disease have multiple potential biological and clinical applications.

2. Objectives I

The main objectives of this part of the thesis can be summarized as follows:

- To explore the mechanisms underlying disease-related genes. We aim to discover and characterize novel interactions between causative and associated genes to breast and colorectal cancers.

- To expand BC and CRC interactomes using a Y2H library approach in order to identify new genes related to each disease.

- Focusing on breast cancer, we aim to integrate our novel interaction data into the data currently available in the literature and apply a module-assisted network approach to predict protein function. The main goal here is to discover the function of proteins whose role in the disease is unknown.

3. Results and Discussion I

3.1 Expansion of Breast Cancer Interactome

3.1.1 Selection of breast cancer causative and associated genes

Our main goal in this part of the project is to decipher the mechanisms underlying disease related genes, and an important factor when performing interaction discovery strategies is the selection of the optimal genes. For this reason we first compiled a comprehensive list of 59 driver genes based on their relevance in breast cancer (BC). Among those, three genes of vital importance for breast cancer tumor development are estrogen receptor, progesterone receptor and ERBB2 (=HER2). We also considered 20 genes that cause genetic predisposition to breast cancer, either with high penetrance (BRCA1, BRCA2 and TP53), intermediate penetrance (e.g. CHK2, ATM, BRIP1, PALB, etc) or low penetrance (e.g. PTEN, STK11, RAD50, etc). Additionally, we included 36 other genes that are described into the literature as causative cancer genes and/or are key genes of relevant cancer pathways (e.g. XRCC3, PHB, RAF1, HRAS, etc). A complete list of the selected driver genes is given in Appendix 1.

On the other hand, the list of susceptibility genes somehow related to BC is endless, but the role of these genes in the development of the disease is barely understood. We mined the literature for genes associated to the disease and genes presenting relevant transcriptomics/genomics aberrations in breast cancer patients. This set comprised 186 candidate genes that we selected based on (1) their altered expression in breast cancer, (2) their loss of heterozygosity (LOH) or mutagenicity, and (3) their involvement in metastasis or related pathological processes.

To maximize the chance of detecting novel relationships and mechanisms, while minimizing already reported interactions, we applied the shortest-path length metric (Materials and Methods) and we filtered out those candidates present in the human interactome that were separated by at least three interactions from the nearest driver. Thus, we discarded 96 genes whose distance to the nearest driver in the network was one (direct interactor) or two (with one intermediate protein), since a molecular link could already be established for these genes. This ensures us that all of the interactions found had never been identified before. As in many cases Y2H screens with trans-membrane domains are dominated by false positives we pruned another 32 genes with known/predicted transmembrane regions (10) from the remaining 90 candidates. Finally, we also excluded genes with unavailable ORFs (22), ending up with 58 candidates. A complete list of the selected associated genes and their features is given in Appendix 2.

3.1.2 Identification of novel BC-related genes through interaction discovery experiments

It has been reported that causative and susceptibility genes in complex diseases, including human cancers, tend to be highly interconnected (Oti and Brunner, 2007). We first checked the interconnectivity between these 59 driver genes to verify whether this observation is also true for BC. To this end, we again computed the minimal distance (i.e. the shortest path length) between a pair of drivers in the frame of the charted human interactome. We observed that the shortest path between driver genes is 2.8 meaning that, on average, we need less than three links (i.e. two intermediate proteins) to physically connect any two gene products within this set. To assess the statistical significance of this figure, we compared this result to a reference distribution consisting of randomly picked sets of 59 proteins in the human interaction space (RND, average shortest path = 4.4). The average distance among BC-related genes was significantly shorter than that of the reference distributions (P-value$_{RND}$ ~ 0), indicating that BC driver genes are indeed more tightly connected than one would expect by chance. We thus sought to exploit the high interconnectivity observed among BC drivers in order to reveal novel direct

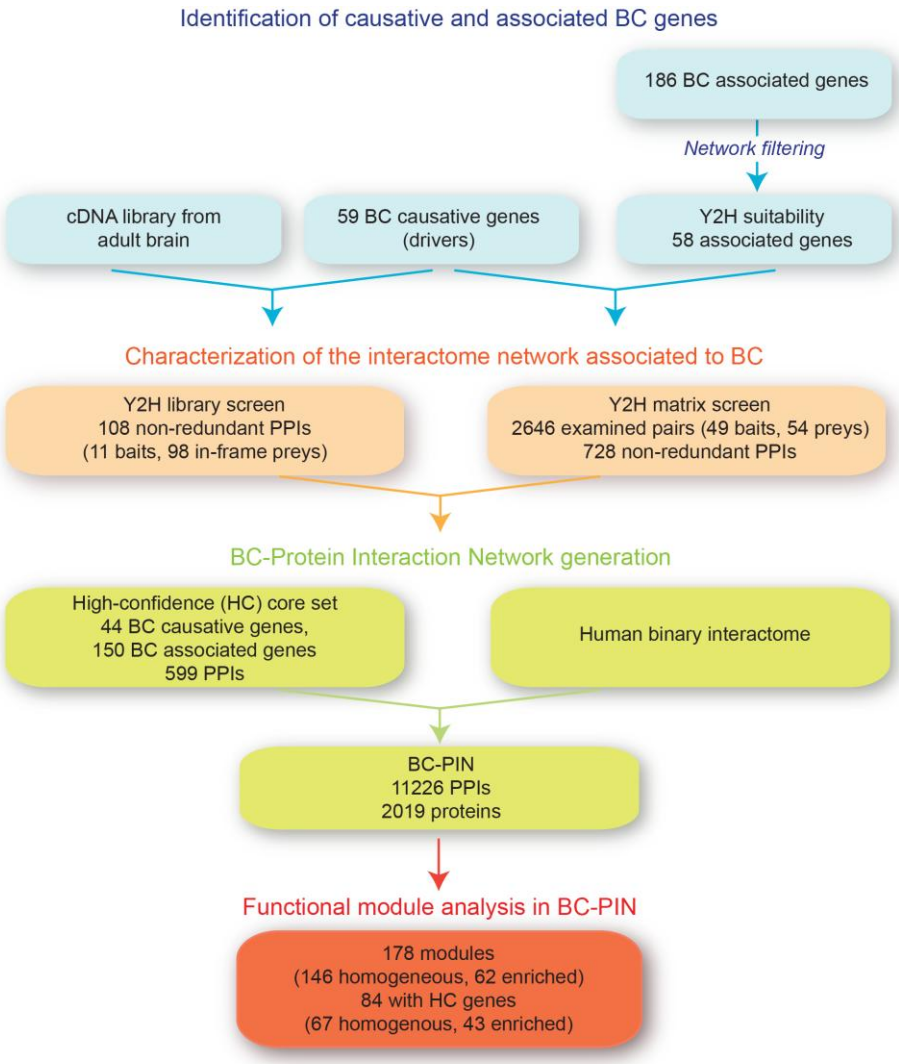


Figure 10. Flow strategy of BC interactome characterization studies.

We followed four major steps to expand and analyze the interactome associated to breast cancer (BC): (1) Identification of causative and associated BC genes. (2) Characterization of interactome network by Y2H screen. (3) Generation of the BC-associated protein interaction network (BC-PIN). (4) Functional module analysis.

relationships between drivers and the set of associated genes selected, which would provide a molecular rationale for the changes observed. The complete flow strategy of BC interactome characterization studies is represented in Figure 10.

For this reason, we performed systematic matrix-based Y2H screens to identify and characterize novel interactions between driver (causative) and candidate (associated) breast cancer genes. We converted 49 drivers into bait and 54 candidates into prey plasmids, respectively (Materials and Methods), since we had to exclude from the analysis 10 drivers and four candidates due to incorrect cloning. We had problems basically when cloning into the pENTR-D-TOPO vector, as after several attempts the sequencing was showing incorrect or absent cloning. This complication was more common with the bigger genes, and some of them such as BRCA1, BRCA2 drivers and KIAA0100 candidate were excluded. In addition, five baits (E2F1, MAPK14, NCOA3, PGR and TP53) were discarded as they showed self-activation in Y2H screens, which indicates that they can activate transcription in the absence of any two-hybrid-interacting partner protein. In order to increment the confidence of our results, each interaction was tested in duplicate and we measured the activation of three different reporter genes: HIS, URA3 and LACZ (see Materials and Methods for further details). Overall, we examined 2,376 pair-wise protein interactions by Y2H and identified 728 non-redundant interactions, which were subsequently scored based on their ability to activate at least two reporter genes or being observed in biological replica screens. We finally selected 491 interactions, which we defined as our matrix-based Y2H high-confidence core (HC) interaction set (Appendix 3), and which represent the first molecular links reported between these 54 associated genes and well understood BC causative pathways.

In order to expand the initial number of interactions, we carried out an alternative identification of novel BC-related genes by Y2H library assays. We selected 11 genes: ER, ERBB2 and those genes involved in early onset breast cancer that were correctly cloned into bait vector and that were not a self-activator bait (genes highlighted in Appendix 1). These 11 baits were screened in five independent replicates against a human cDNA library. We identified 312 interactions, from which 118 contained the downstream gene in frame with the

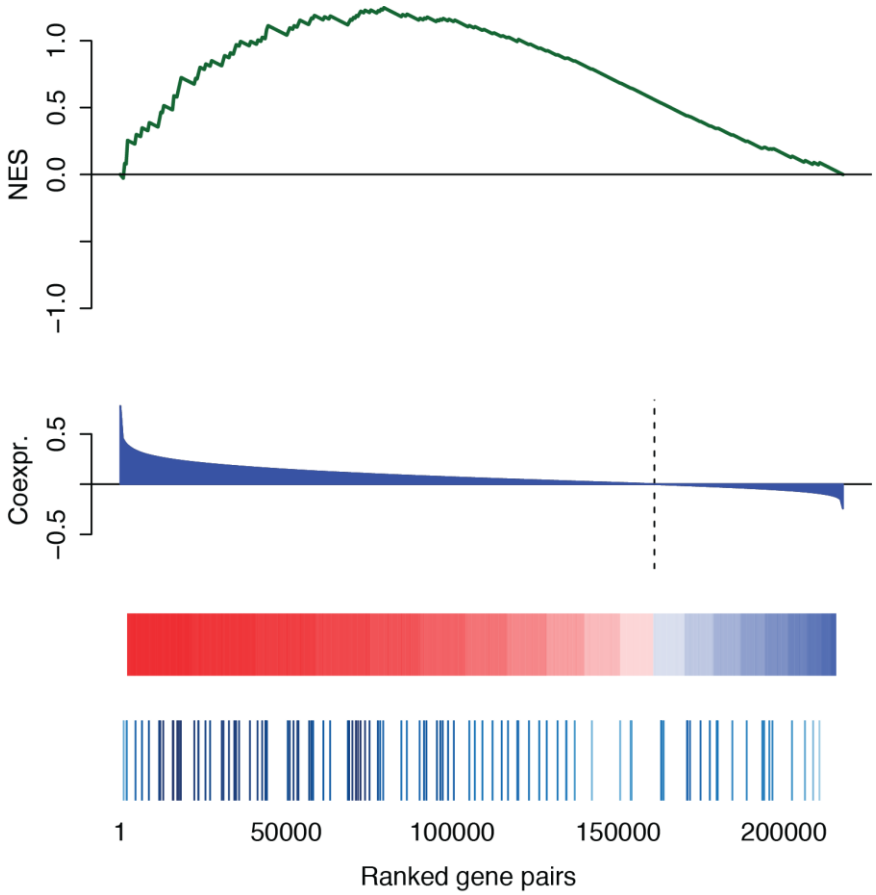


Figure 11. Gene co-expression analysis of Y2H library screen interactions.

Similar to GSEA (Subramanian et al., 2005), the running sum is the result of the weighted counting of gene pairs of interest through a ranked list involving all possible pairings of drivers/causing genes against a comprehensive gene catalogue. In the upper panel, the running sum is measured as a normalized enrichment score (NES). The weighting is done according to correlation values obtained from COXPRESdb (Obayashi et al., 2013) (second and third panels). In the bottom panel, gene pairs corresponding to our interactions are highlighted.

prey activation domain. These interactions were re-tested by matrix Y2H experiments to validate their interaction specificity, ending up with 108 new HC interactions for the 11 baits assayed (Appendix 4). Notably, 76 of the 99 newly identified proteins had never been linked to breast cancer before, according to Intogen database (Gundem et al., 2010). Nevertheless, these genes are significantly co-expressed in human tissues with their BC driver interaction

partners (P-value = 0.024; see Materials and Methods and Figure 11), and significantly overrepresented in tumor somatic mutations (P-value < 0.016), strengthening their potential relevance in BC and validating our strategy.

Based on the outcome derived from both Y2H screens, we generated a definitive HC protein-protein interaction network comprising 599 non-redundant interactions, involving 44 drivers and 150 candidates (Figure 12).

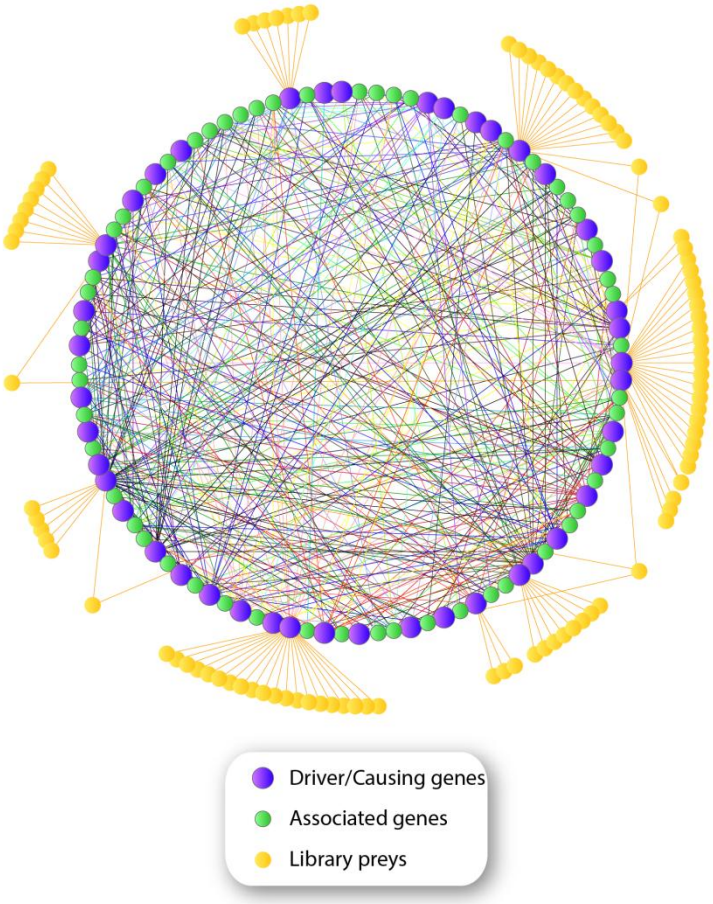
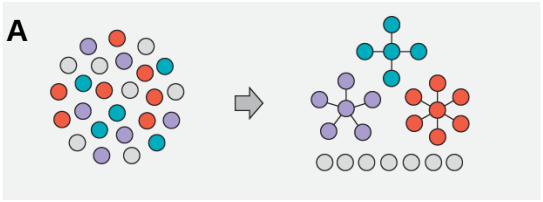


Figure 12. High-confidence interaction network.

Diagram showing the 599 novel interactions reported from our Y2H screen. Drivers are depicted in purple, candidates in green and library-interacting preys are highlighted in yellow.

Strikingly, we only identified one common interaction to both matrix and library screens (ATM-WHSC1L1). Furthermore, we also detected an interaction between two drivers in the Y2H library assays (STK11-TSG101). Of note, only one interaction, ESR1-ATAD2, has been reported previously (Zou et al., 2007), indicating that almost all of our HC interactions are novel, as expected. The low overlap between different networks is a well-known effect, and it is mainly attributed to the limited sampling of the interactome space (Venkatesan et al., 2009).



B

P-value 0.5 0.25 0.1 0.05 $<10^{-4}$

AKT1	APC	ATM	AURKA
BAG4	BCAR3	BRMS1	CASP8
CCND1	CDH1	CDKN2A	CDKN2C
CHEK2	ERBB2	ESR1	ESR2
FBXW7	FGFR2	FGFR4	HMMR
HRAS	IGF1R	KRAS	LSP1
NBN	NOTCH2	PALB2	PARP1
PAX2	PHB	PIK3CA	PPM1D
PTEN	PTPN1	PTPRJ	RAD51
RAF1	RB1CC1	SMAD4	STK11
TGFB1	TSG101	WT1	XRCC3

Figure 13. Functional coherence of interaction network

(A) Conceptual framework to exploit networks for the understanding of biological processes. We relate independent biological annotations through PPIs. In this scheme functions are depicted by different colors. **(B)** The biological function coherence of every driver with its interactors is color-graded, from an overall trend to significant coherence (bottom panel).

To evaluate whether the discovered interactions could really provide mechanistic details about the relationships between driver and candidate genes, we assessed the functional coherence of our HC interaction network by checking if each driver and its interacting candidates are generally involved in similar biological processes, as assigned by Gene Ontology terms. Very interestingly, the network shows a high degree of biological process coherence (pairwise semantic similarity mean value = 0.347, P-value < 0.0001) (see Materials and Methods). The same analysis with driver direct interactors from canonical pathways also shows a high coherence (pairwise semantic similarity mean value = 0.627, P-value < 0.0001), as expected. These findings indicate that our network-based approach is indeed a robust inference tool to gain insight into the underlying mechanisms of those proteins with previously unknown roles in BC, as well as for a better understanding of the regulation and interrelationship between different proteins of complex biological systems (Figure 13).

All the protein interactions reported here have been submitted through MINT (Ceol et al., 2010) to the International Molecular Exchange (IMEx) Consortium (http://www.imexconsortium.org) and assigned the identifier IM-21668.

3.1.3 Interactome network associated to BC

To contextualize the 599 novel BC-related interactions between the 194 proteins that our study has revealed, we integrated them with the human interaction data currently available in the literature to build a comprehensive interactome associated with BC. We retrieved from the databases all the proteins identified as direct interactors of the group of drivers considered in our study and merged them with our HC set of interactions. Additionally, we further extended this initial network to the next level (i.e. we included all the direct interactors to the initial set and the interconnections among them), obtaining a network of 11,226 interactions among 2,019 proteins, designated as BC-protein interaction network (BC-PIN).

We next studied the structure of the BC-PIN to detect the presence of potential functional modules, defined as groups of proteins that are densely interconnected and that are functionally homogenous (i.e. functional

annotation shared by the maximum number of module proteins). To identify these modules, we used the MCL algorithm (Enright et al., 2002), since it has proved to be more robust and tolerant to noise than other modules detection methods (Vlasblom and Wodak, 2009). With this procedure, we identified 178 modules in the BC-PIN, of which 146 showed a high degree of functional homogeneity, roughly containing 64% of the proteins in the network. Additionally, we found that 62 of them were significantly enriched for one or more GO biological process annotation. If we look for the positioning of the drivers and their interactors in our HC set in the BC-PIN, we find that they have been grouped into 84 distinct modules (72% and 58%, respectively), 67 of which are homogenous for, at least, one GO annotation. When we reanalyzed these functional modules excluding our 599 HC interactions, we observed that the number of modules falls from 178 to 112, and the number of homogeneous and enriched groups also decreases from 55 to 26 modules, respectively. This supports the idea that our study charts some unexplored areas of the BC-associated interactome and significantly increases the coherence of known modules with more connections. All the BC-PIN modularity data is reported in Appendix 5.

Eventually, we exploited the modular structure of the BC-PIN to extract relevant sub-networks representing the most frequent homogeneous functions in the interactome. We merged the modules with similar functions (Materials and Methods) to build sub-networks, each of them containing all the modules that share a biological function, and then we added the directly interacting candidates and those linking at least two modules from our HC set. The resulting four main sub-networks, which are represented in Figure 14, are related to signal transduction (7 modules, 157 proteins and 221 interactions), DNA damage repair (4 modules, 87 proteins and 167 interactions), protein phosphorylation (4 modules, 99 proteins and 114 interactions) and apoptosis (2 modules, 69 proteins and 109 interactions).

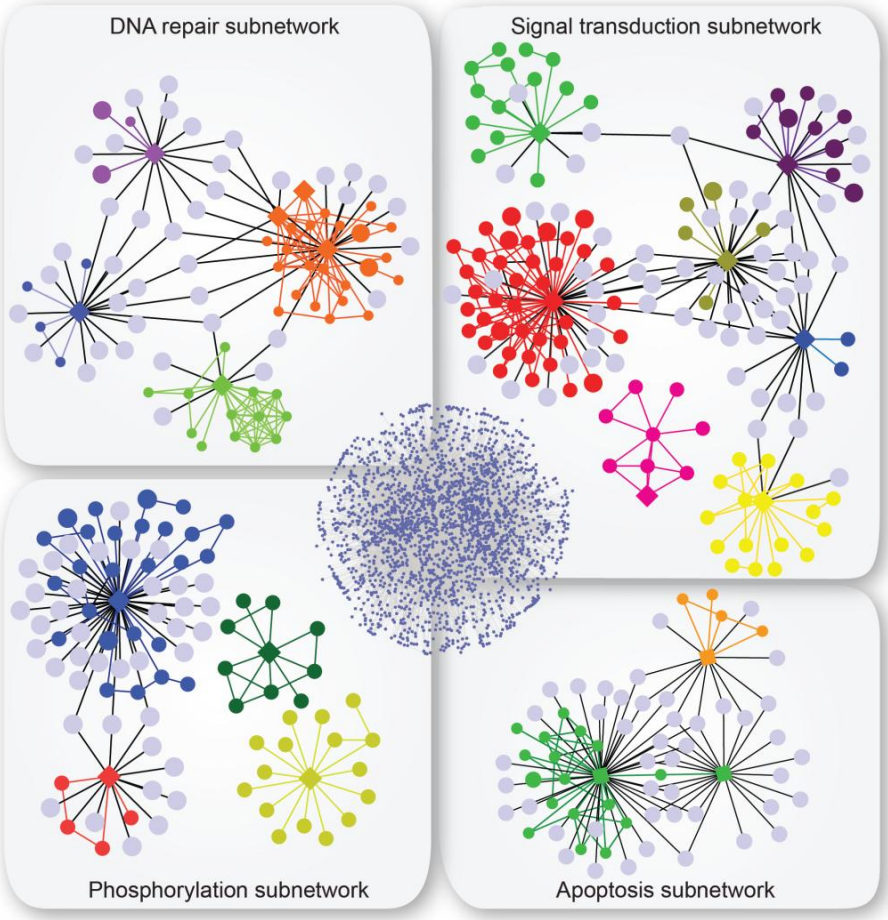


Figure 14. BC-PIN subnetworks

Diagrams showing the four relevant sub-networks based on the most frequent homogeneous functions in the BC-PIN. The clusters were merged by similar functions to build sub-networks, each of them containing all the modules that share a biological function. The resulting four main sub-networks are related to signal transduction (7 clusters), DNA damage repair (4 clusters), protein phosphorylation (4 clusters) and apoptosis (2 clusters).

3.2 Expansion of Colorectal Cancer Interactome

3.2.1 Selection of CRC causative and associated genes

As we mentioned in Section 1.3.4, it has been described that, in complex diseases, causative and susceptibility genes tend to be highly interconnected. Thus, we performed a similar approach as described for BC interactome charting (see Section 3.1) but directed to CRC interactome. We again performed a systematic Y2H screening to reveal novel direct relationships between CRC causative genes, namely drivers, and a set of CRC associated genes, namely candidates. We propose that, by discovering interactions between drivers and candidate genes, we can help unraveling the role of the latter in the CRC phenotype and provide mechanistic details for such relationships.

The selection of CRC driver genes was conducted similarly to the BC driver genes' selection. We compiled a list of 45 driver genes. Eleven of these drivers were manually selected for being well-established and characterized genes in CRC development (Markowitz and Bertagnolli, 2009) and/or they are key genes in the main signaling pathways involved in colorectal cancer syndromes (Sancho et al., 2004). In addition, we included 24 genes whose defects can strongly contribute to an inherited predisposition to CRC, according to OMIMs database (McKusick, 2007). Mutations in one single gene out of those 24, result in a marked predisposition to colorectal cancer in any of the two distinct syndromes: familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC). Finally, by examining COSMIC's database (Bamford et al., 2004; Forbes et al., 2010), we selected another 10 drivers, which are genes whose somatic alterations are known to be crucial in sporadic CRCs. A complete list of the selected driver genes is given in Appendix 6.

There are hundreds of CRC associated genes: genes that have been related CRC by several assays such as expression arrays or gene linkage analyses and genome-wide association studies, but whose role in the disease is barely understood. To reduce the number of candidate genes and maximize the biological relevance of the discovered interactions, we selected 45 susceptibility

genes for the interaction discovery experiments. The main difference among BC and CRC interactome characterization studies that we performed resides in the criteria applied to select the associated genes tested by Y2H. In the case of CRC, we employed the following criteria: (1) First we prioritized genes with coordinated expression changes with CRC drivers across a compendium of normal tissues and cell types. We estimated co-expression in terms of correlation coefficients computed using an Expectation-Maximization EM algorithm (Dempster et al., 1977). Besides, in all the cases, we forced a co-expression in the colon and rectum related tissues to obtain the most relevant correlation for CRC (see Materials and Methods). This procedure identified 17 candidate genes that did significantly co-express with the known CRC-causative genes. (2) Besides, we considered genes within chromosomal loci identified by gene linkage analyses and genome-wide association studies as related to CRC. In particular, we included the region 9q22.32-31.1, which contains a susceptibility locus (CRCS9) involved in the development of known hereditary colorectal cancer syndromes, familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC) (Kemp et al., 2006; Skoglund et al., 2006; Wiesner et al., 2003). It is known that these loci are linked to CRC, but the specific genes responsible for this linkage remain to be discovered. Accordingly, we selected 28 additional genes from this locus potentially implicated in CRC disease mechanisms. Eventually, we obtained a final list containing 45 CRC susceptibility genes, which is given in Appendix 7.

3.2.2 Identification of novel CRC-related genes through interaction discovery experiments

We then performed systematic matrix-based yeast two-hybrid (Y2H) screens to identify novel interactions between driver and susceptibility CRC genes. We converted 45 drivers into bait and 45 associated genes into prey plasmids, respectively (see Materials and Methods). From those, three baits were discarded as they showed self-activation in Y2H screens in presence of empty prey clones (BUB1B, CTNNB1 and TP53). Overall, we examined a matrix of 1,890 protein pairs by Y2H and identified 1,029 non-redundant interactions, which were also subsequently scored based on their ability to activate at least two reporter genes or on being observed in the two biological replica screens.

We finally selected 595 interactions, which we defined as our matrix-based Y2H high-confidence core (HC) interaction set (Appendix 8), and it represents the first molecular links reported between these 45 susceptibility genes and well-understood CRC driver genes and pathways.

As previously performed for BC interactome, we also carried out an alternative identification of novel CRC-related genes by Y2H library assays. As baits, we selected six genes based on gene expression profiles from normal mucosa, colorectal adenoma and adenocarcinoma (Sabates-Bellver et al., 2007). In particular, we focused our study on the driver genes AXIN2, DLC1 and PDGFRL, which are highly up-regulated in CRC compared to adenoma (benign tumor) or normal mucosa, and the candidates C9orf30, SFRP4 and SFRP2, up-regulated in colorectal tumors compared to both normal mucosa and benign tumors. Hence, these alterations suggested that these might be specific genes for CRC tumorigenesis. After converting these genes into bait plasmids, we screened them against a human cDNA library through yeast two-hybrid (Y2H) library assays. We carried out 30 Y2H screens against an adult brain cDNA prey library (5 independent replicates for each of the six baits), which yielded 105 interactions between the six baits and 74 distinct cDNA clones or preys. DNA sequence verification and a systematic BLAST search showed that 33 of the isolated potential interactors (i.e. preys) contained the downstream gene in frame with GAL4 activation domain, while the remaining clones showed out-of-frame sequences or sequences from non-coding regions, which were discarded. These 33 potential interactors were re-tested by matrix Y2H experiments to validate their interaction specificity. We could validate 27 of them, involving 20 novel proteins, which we then considered as high-confidence interactions and further included in the CRC interactome (Appendix 9). Most of the identified preys interacted with a single bait, and only PFKM, RANGAP1 and TMEM129 interacted with two or more baits.

Based on the outcome derived from both Y2H screens, we generated a definitive HC protein-protein interaction network comprising 622 non-redundant interactions, involving 42 of the selected CRC driver genes (Figure 15). Interestingly, we did not identify common protein-protein interactions in both matrix and library screens, which reveals the advantage of performing pair-wise and pool screens in parallel. It is worth mentioning that only six

interactions have been reported since we started the study (Mosca et al., 2013), which indicates that almost all of our HC interactions are novel, as expected. The low overlap between different networks is a well-known effect, and it is mainly attributed to the limited sampling of the interactome space (Venkatesan et al., 2009). In our case, this is particularly pronounced since we chose some of our candidate genes to maximize the number of novel interactions added to the CRC associated network and of which little

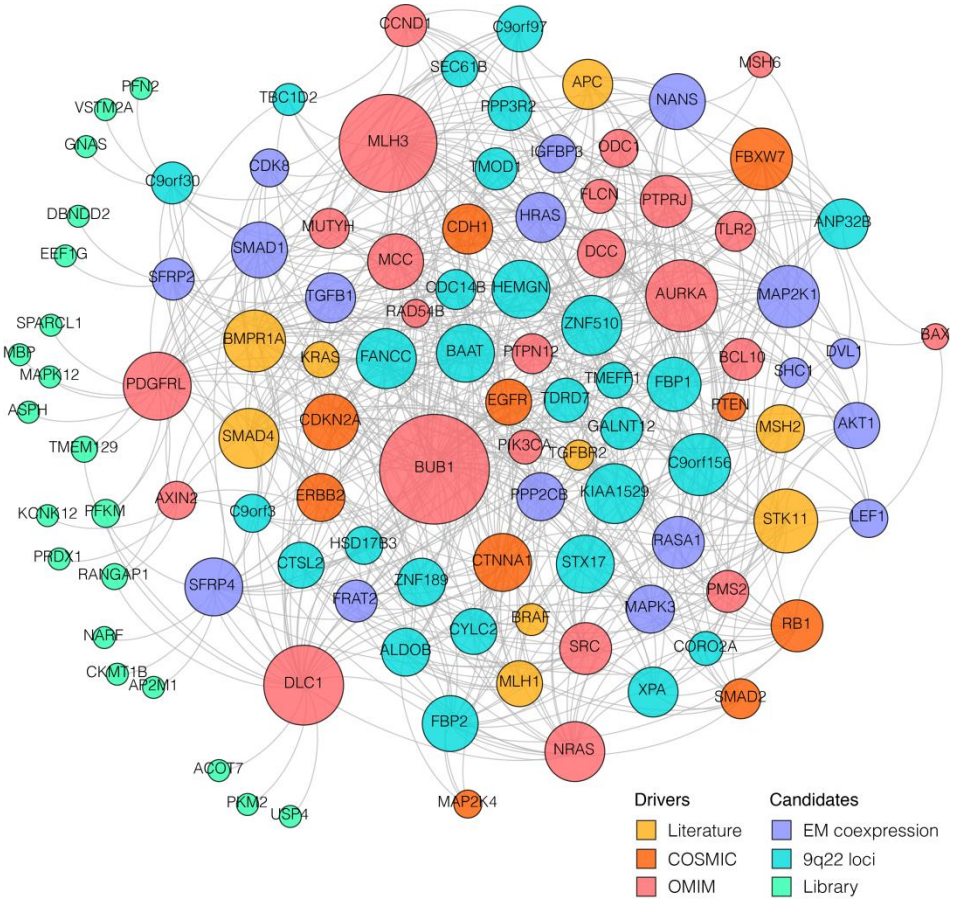


Figure 15. CRC interaction network.

Visual representation of the relationships between CRC drivers and HC interactors discovered. Drivers are depicted as yellow (literature curated), orange (COSMIC database) and red (OMIM database) nodes, whereas CRC associated genes are depicted as purple (co-expression correlation), blue (CRCS9 loci) and green (library interactors) nodes. The size of each node is proportional to its number of interactions.

interaction information was known (i.e. genes in the susceptibility region CRCS9 with no direct proof of their implication in CRC).

To evaluate whether the discovered interactions could really provide mechanistic details about the relationships between driver and candidate genes, we assessed the functional coherence of our HC interaction network. As previously described for BC interactions (see Section 3.1.2), we checked if each driver and its interacting candidates are related to similar biological processes, as assigned by Gene Ontology terms. Interestingly, the network shows a high degree of biological process coherence (pairwise semantic similarity mean value = 0.321, P-value < 0.0001) (Figure 16). The same analysis with driver direct interactors from canonical pathways also shows a high coherence (pairwise semantic similarity mean value = 0.610, P-value < 0.0001), as expected. These findings indicate that our network-based strategy is useful to gain insight into the underlying mechanisms of those proteins with previously unknown roles in CRC, as well as to have a better understanding of the regulation and interrelationship between different proteins of complex biological systems

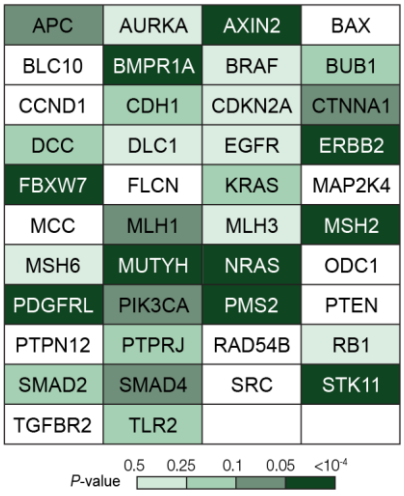


Figure 16. Functional coherence of CRC interaction network.

The biological function coherence of every CRC driver with its interactors is color-graded, from a global tendency to significant coherence.

3.3 Identification of novel BC-genes associated with the DNA Damage Response

3.3.1 DNA damage repair sub-network

The DNA damage repair emerged as one of the most interesting interaction sub-networks (Section 3.1.3; Figure 17). Deficiencies in DNA damage signalling are both known to play causative roles in BC and are actively exploited for therapeutic protocols in cancer treatment (Lord and Ashworth, 2012). We thus decided to focus on this sub-network to estimate the precision of our interaction set and illustrate the power of our approach to generate novel mechanistic hypothesis.

The clusters within the DNA damage repair sub-network contain six driver genes related to DNA repair, four candidates from our matrix-based HC set, and two HC library-preys (Figure 17). In particular, the largest cluster contains 25 proteins, with three drivers (ATM, NBS1/NBN and RAD50) that are related to the DNA damage response through MRE11-RAD50-NBS1 (MRN) complex sensing of DNA breaks and activation of the ATM kinase (Stracker and Petrini, 2011). Two more clusters involving RAD51 (containing five proteins) and XRCC3 (containing four proteins), are also related to DNA repair by homologous recombination (HRR), one of the two major double-strand break repair pathways (Liu et al., 2007). Finally, the PARP1 cluster (14 proteins) is involved in base excision repair (BER) and is a potential chemotherapeutic target in HRR deficient tumors, such as those lacking BRCA1 or BRCA2 (Rouleau et al., 2010).

Although it is well-documented that different interaction discovery techniques are able to identify interactions of different nature (i.e. binary/multimeric, transient/dedicated, etc) (Venkatesan et al., 2009), we sought to validate some of our interactions derived from Y2H screens with alternative strategies. To this end, we selected a random subset of the HC interactions containing DNA damage repair-related genes to validate them in mammalian cells by using complementary techniques, expecting that the results obtained would be representative of the entire HC interaction set (Figure 18A). We first tested 14 interactions by co-IP binding experiments in HEK293 cells (see Materials and

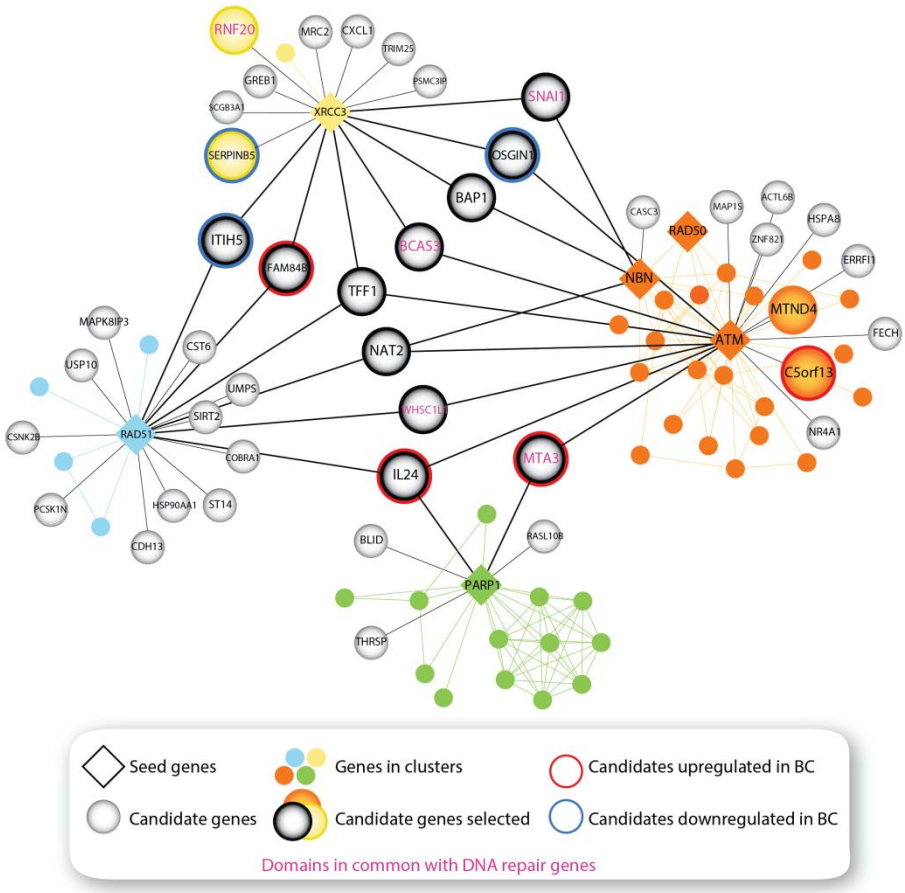


Figure 17. DNA repair-related sub-network.

Interactions between DNA damage drivers and HC interactors. Drivers are depicted as rhombus and HC interactors as circular nodes. The four homologous functional clusters are differently colored. Genes with altered expression in BC tissues (Oncomine™ database) are highlighted in red (up-regulated) and blue (down-regulated) and those genes containing common domains with DNA repair proteins, in purple.

Methods), and could confirm 11 of them, corresponding to a 79% of validation rate (Figure 18B). We also studied the subcellular distribution pattern of several DNA damage repair-related genes by their overexpression in U2OS cells and subsequent observation under a confocal microscope. We tested nine interacting pairs and we observed that four of them shared predominantly

nuclear location (CHK2-FAM84B, RAD51-FAM84B, RAD51-NAT2 and RAD51-WHSC1L1) (Figure 18C). Taken together, these results strongly support our Y2H HC interaction set.

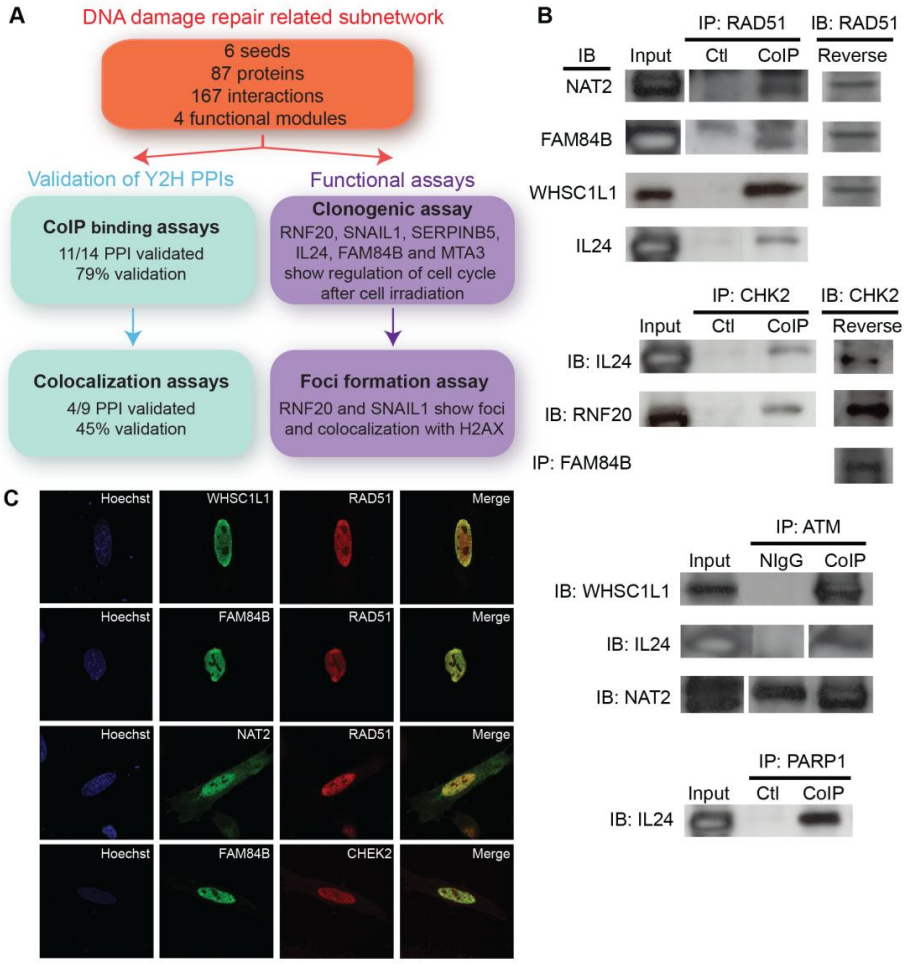


Figure 18. Validation of Y2H interactions by downstream binding assays.

(A) Schematic diagram showing the experimental outline to assess the DNA damage and repair subnetwork coherence. (**B)** Interactions examined by co-IPs. Fourteen HC interactions involving four drivers of the subnetwork were tested. Eleven of them were confirmed (79% validation rate). Input: cell lysate, loading control. IP: immunoprecipitated protein. IB: immunoblotted protein. Ctl: Empty vectors, negative controls. NIgG: non-immune rabbit or mouse immunoglobulins, used to monitor non-specific IPs. **(C)** In vivo co-localization of interacting partners by immunofluorescence microscopy. RAD51-WHSC1L1, RAD51-FAM84B, RAD51-NAT2 and CHK2-FAM84B share nuclear co-localization in U2OS cells. Merged images show co-localized regions in yellow.

To further investigate the relationships of genes present in our sub-network to DNA damage repair, we focused on 15 candidate genes that were either present in DNA repair functional clusters or linking two or more DNA repair modules (Figure 17). Interestingly, five of these genes (BCAS3, MTA3, RNF20, SNAI1 and WHSC1L1) share common domains with many DNA damage response enzymes (Table 3), but only RNF20 has been described as a modulator of the DNA damage response (Nakamura et al., 2011). Furthermore, seven of the remaining genes are up to ten-fold either down-regulated (ITIH5, OSGIN1 and SERPINB5) or overexpressed (C5orf13, FAM84B, IL24 and TFF1) in BC compared to normal breast tissue (Rhodes et al., 2007), but a potential relationship with DNA damage response has never been established.

Gene symbol	SMART domains	Pfam-A domains
BAP1	-	Peptidase_C12
BCAS3	WD40	WD40
C5orf13	-	Alveol-reg_P311
FAM84B	-	NC
IL24	-	-
ITIH5	VWA; VIT	VWA; ITI_HC_C; VIT
MT-ND4	-	Oxidored_q1; Oxidored_q5_N
MTA3	SANT; ZnF_GATA; BAH	BAH; GATA; ELM2
NAT2	-	Acetyltransf_2
OSGIN1	-	Pyr_redox_2
RNF20	RING	zf-C3HC4
SERPINB5	Serpin	Serpin
SNAI1	ZnF_C2H2	zf-C2H2
TFF1	PD	Trefoil
WHSC1L1	SET; AWS; PHD; PWWP; PostSET	PWWP; SET

Table 3. Functional domain annotations of selected candidate genes.

We checked if any of the candidates had at least one Pfam-A or SMART domain found in any gene of the two reference gene sets containing DNA repair genes. The domains highlighted in red are found in genes belonging to the DNA repair set, whereas the ones highlighted in blue are present in genes of the DNA damage response set.

3.3.2 Initial functional validation of the BC-genes associated to DNA damage response

As a first approach to study whether our candidate genes are indeed related to the DNA damage response, we carried out clonogenic survival assays to assess the sensitivity to radiation after overexpression of our genes of interest (see Materials and Methods). We successfully overexpressed 11 genes by cDNA plasmid transfection in HEK293T cells (BAP1, BCAS3, FAM84B, IL24, MTA3, NAT2, OSGIN1, RNF20, SERPINB5, SNAI1 and WHSC1L1), while we had to discard four genes after unsuccessful cloning and transfecting attempts (ITIH5, C5orf13, MT-ND4 and TFF1).

To establish the reference lines of cell survival upon ionizing radiation (IR) treatment, we used as a negative control cells transfected with the empty vehicle vector and, as a positive control, cells overexpressing RAD51, a well-established protein involved in HRR and cell proliferation (Li et al., 2000). Of our candidate genes, three of them showed a significant increase in cell survival upon irradiation (P-value < 0.01), as we observed by the radiation survival curve (Figure 19A). SERPINB5 (P-value = 6.00×10^{-5}) and RNF20 (P-value = 7.86×10^{-4}) showed increased cell survival and growth particularly at high IR doses (2-3 fold higher survival as compared to control at six Gy irradiation dose) whereas SNAI1 (P-value = 6.03×10^{-3}) showed a continuously increased colony formation at two, four and six Greys (Gy), in agreement with previous findings of increased cell resistance to DNA damage-induced apoptosis after aberrant expression (Kajita et al., 2004) (Figure 20; Table 4). Conversely, we observed no change in survival after irradiation in cells overexpressing FAM84B, IL24, WHSC1L1, MTA3, BCAS3, NAT2, BAP1 or OSGIN1.

Similarly, we performed clonogenic survival assays after silencing the expression of our genes of interest using siRNA. HEK293T cells were transiently transfected with siRNA against the target genes, followed by ionizing radiation treatment (see Materials and Methods). From the initial list of 15 genes, we successfully depleted the expression of seven genes, namely RNF20, WHSC1L1, SERPINB5, BCAS3, MTA3, IL24 and FAM84B. Luciferase-targeted siRNA was used as a negative control and all the results were normalized to non-treated cells.

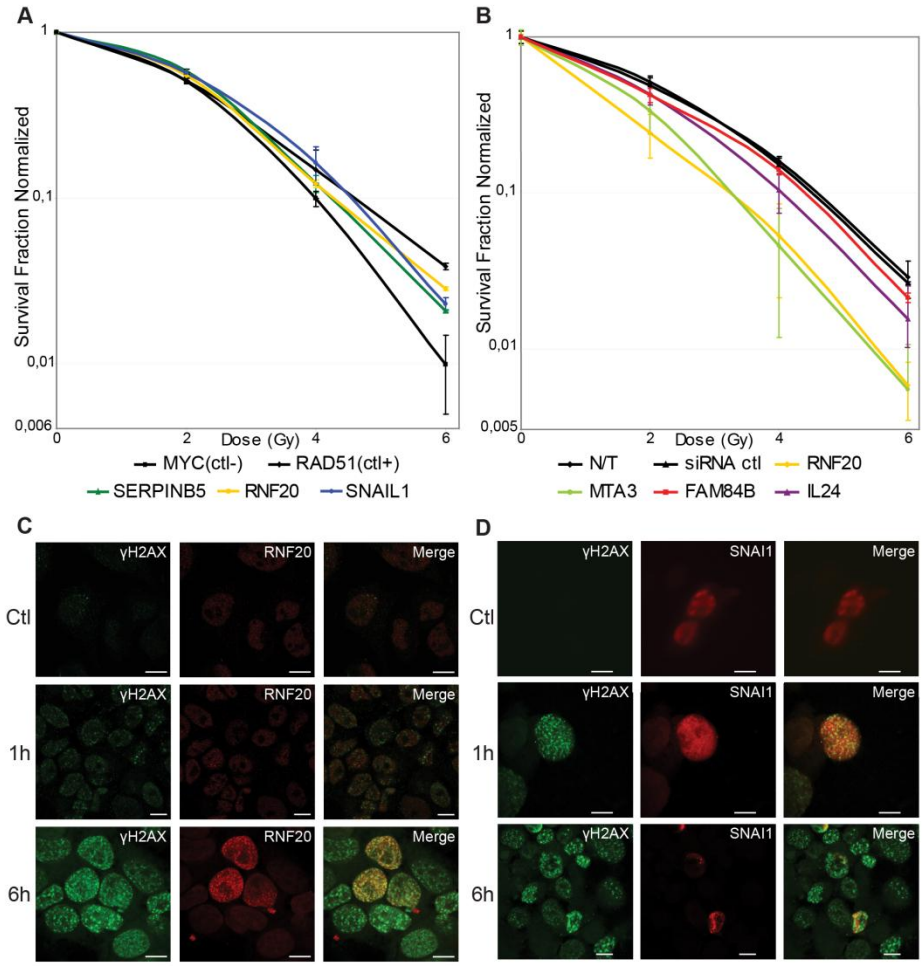


Figure 19. Functional assays.

Clonogenic cell survival assays after **(A)** overexpression or **(B)** silencing of selected candidate genes potentially involved in DNA damage in U2OS irradiated cells. Six genes (*RNF20*, *SERPINB5*, *SNAIL1*, *FAM84B*, *MTA3*, *IL24*) show modulation of cell growth after DNA damage. Foci formation assays after induced DNA damage show co-localization of (**C**) RNF20 and (**D**) SNAIL1 with γH2AX, indicating they are recruited at the DNA DSBs. γH2AX foci formation was detected by using a rabbit monoclonal anti-γH2AX antibody. Alexa Fluor 488-labeled goat anti-rabbit IgG (green) and Alexa Fluor 568-labeled goat anti-mouse IgG (red) secondary antibodies were used. Scale bars represent one μm. Data are represented as mean ± SD.

After irradiation with a dose of two-six Gy, four out of the seven genes (RNF20, MTA3, IL24 and FAM84B) showed higher radiation sensitivity, exhibited by diminished colony proliferation relative to the negative control siRNA (Figure 19B). Although we clearly observed a significant impact on cell growth and survival in the four cases, MTA3 and RNF20 showed the most dramatic effects, decreasing colony proliferation by 70-75% (P-value = $3.00x10^{-6}$ and $2.18x10^{-5}$, respectively), whereas IL24 (P-value = $2.14x10^{-4}$) and FAM84B (P-value = $4.53x10^{-4}$) depletion only decreased survival to 25-30% relative to the control (Figure 21; Table 4B).

Consistent with previous data linking it to double strand break repair, RNF20 displayed coherent IR survival patterns between overexpression and silencing assays (Moyal et al., 2011). Conversely, depletion of SERPINB5 did not show a significant decrease in cell survival, while it increased radioprotection when overexpressed. It is worth mentioning that SERPINB5, commonly known as MASPIN, also displays an heterogeneous behavior in clinical samples, being highly underexpressed in all BC types, and overexpressed in other cancer types such as colorectal, lung or pancreas (Rhodes et al., 2007).

The depletion of the MTA3 gene also induced an enhanced radiosensitivity relative to control in HEK cells, in consensus with its role as a transcriptional repressor of SNAI1, a major regulator of the epithelial to mesenchymal transition (EMT) that is critical for invasive growth of breast cancers (Fujita et al., 2003).

Intriguingly, IL24 and FAM84B enhanced cell radiosensitivity when inhibited but did not show a significant effect when overexpressed (P-value < 0.01). Since IL24 has been shown to cause antiproliferative and cytotoxic effects in a variety of tumor cells, but not in non-transformed cells (Su et al., 2003; Yacoub et al., 2003), our findings suggest that they might be indeed under tight regulation to contribute to terminal cell differentiation, leading to higher sensitivity when expression is downregulated.

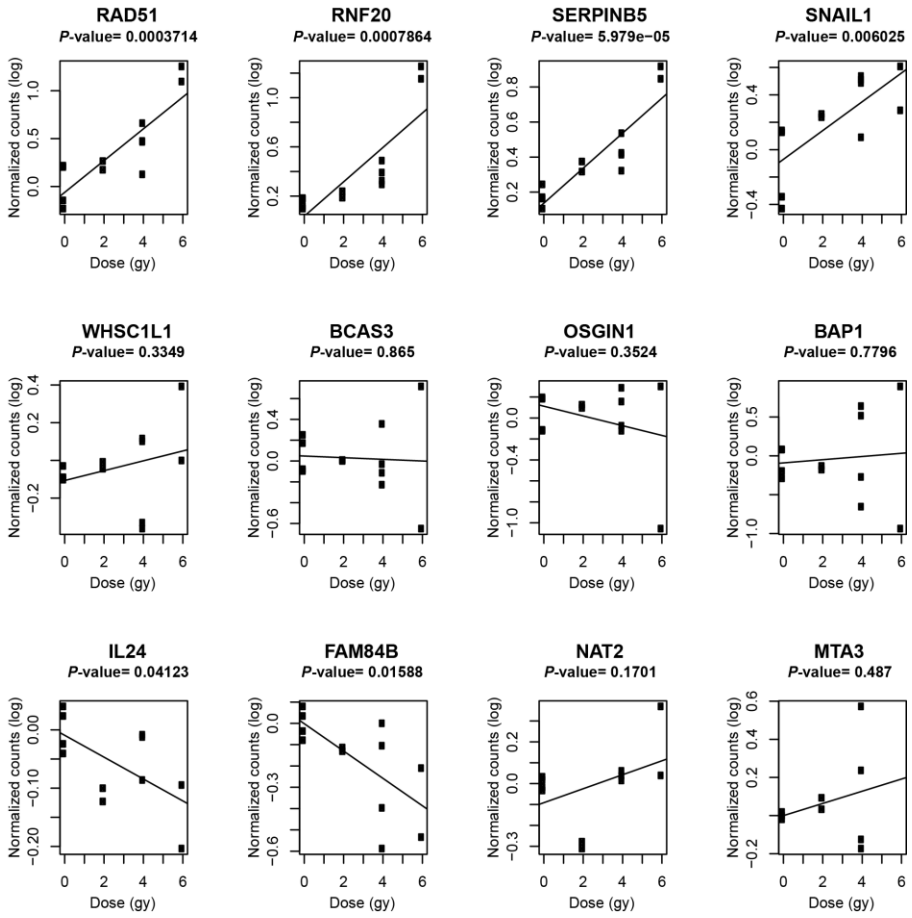


Figure 20. Statistical analysis of clonogenic assays (overexpression).

Plot diagrams of normalized colony counts of each gene when overexpressed. Normalized counts are depicted as black spots; trend line and P-value are also shown.

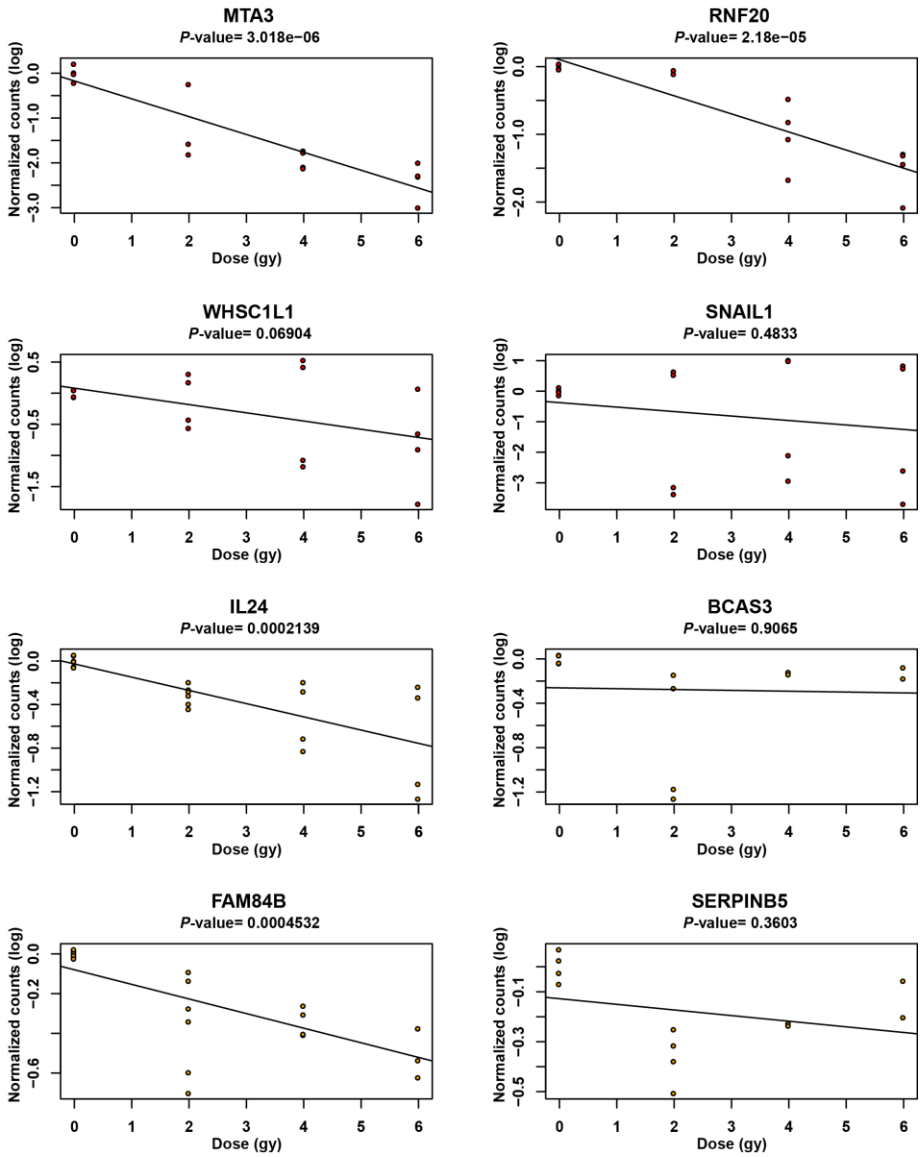


Figure 21. Statistical analysis of clonogenic assays (silencing).

Plot diagrams of normalized colony counts of each gene when silenced. Normalized counts are depicted as black spots; trend line and P-value are also shown.

(A) Gene	0 Gy	2 Gy	4 Gy	6 Gy	Trend *P*-values
SERPINB5	**0,0048**	**0,0068**	0,0601	**0,0015**	**0,000060**
RAD51	0,9110	0,0405	0,0841	**0,0043**	**0,000371**
RNF20	**0,0075**	0,0160	0,0867	**0,0017**	**0,000786**
SNAIL1	0,3448	**0,0029**	0,1006	0,1070	**0,006025**
FAM84B	0,0872	0,0188	0,5091	0,2377	0,015879
IL24	**0,0008**	**0,0082**	0,0280	0,1905	0,041229
NAT2	0,0449	0,1801	0,0608	0,3582	0,170109
WHSC1L1	0,0274	0,3686	0,2974	0,8208	0,334886
OSGIN1	0,6333	0,0340	0,4689	0,7420	0,352353
MTA3	0,0181	0,0260	0,9659	NA	0,487002
BAP1	0,3547	0,0300	0,8253	0,9860	0,779590
BCAS3	0,4218	0,7292	0,9900	0,9768	0,864994

(B) Gene	0 Gy	2 Gy	4 Gy	6 Gy	Trend *P*-Values
MTA3	0,76823	0,56859	**0,06406**	**0,01759**	**0,0000030**
RNF20	0,04095	0,45338	**0,00079**	**0,00528**	**0,0000218**
IL24	0,40330	**0,00002**	**0,01687**	0,10477	**0,0002139**
FAM84B	0,47364	0,02618	0,00689	0,08373	**0,0004532**
WHSC1L1	0,55113	0,59006	0,75767	0,71010	0,0690363
SERPINB5	0,04653	**0,00012**	0,52076	0,90460	0,3603025
SNAIL1	0,58053	0,44852	0,54748	0,35846	0,4833310
BCAS3	**0,00645**	0,01811	0,83090	0,91834	0,9064926

Table 4. Values from the statistical analysis of clonogenic assays.

Trend-test analysis results based on the collected colony counts after different IR dose (in duplicates). Summary tables showing the gene overexpression **(A)** and silencing **(B)** statistical P-values at different irradiation doses (Gy). Significant values are highlighted in bold. Dashed line delimits the significant threshold at P-value <0.01.

To complement the clonogenic assays after irradiation, we also assessed the localization of our candidate gene products to DNA double-strand breaks (DSBs), the predominant cytotoxic lesion caused by radiation (Hoeijmakers, 2009; Reinhardt and Schumacher, 2012). Within seconds after the generation of DSBs, subnuclear foci known as IR-induced foci (IRIF) are assembled at the break sites. These IRIF arise from chromatin remodeling and orchestrated recruitment of various DNA damage response (DDR) proteins, which are important for mediating the signaling and repair of the damaged DNA, as well as cell cycle checkpoint activation or apoptosis. Phosphorylation of the histone variant H2AX at Ser139 (γH2AX) is among the earliest signaling modifications (Breitkreutz et al., 2008). For this reason, we determined whether the 15 candidate genes potentially involved in DNA damage response localized to IRIF (see Materials and Methods) following 10Gy of IR. U2OS cells overexpressing the candidate genes were analyzed for IRIF formation and colocalization with γ-H2AX 1h and 6h post-irradiation (Bonner et al., 2008). Both RNF20 and SNAI1 re-localize into IRIF upon the induction of DNA damage by IR (Figure 19C-D), providing further support that the recruitment of these proteins to IRIF may be indicative of a role DSB repair or signaling.

3.3.3 Unveiling additional roles of RNF20 in DNA repair

The E3 ubiquitin ligase RNF20 has been found to be hypermethylated in BC (Shema et al., 2008), and our findings point towards a potential role in DNA repair. In the BC-PIN, it is included in a cluster homogeneous for DNA recombination and DNA repair (Figure 17), containing also the XRCC3 and RAD51C proteins. In addition, it contains domains that are common in DNA repair proteins (Table 3). Furthermore, we found that RNF20 promotes radioresistance when overexpressed and cell death when silenced after irradiation (Figure 19A-B). And finally, we detected foci formation and co-localization with yH2AX, demonstrating RNF20 recruitment to DSB sites (Figure 19C).

Our observations are consistent with recent studies that have demonstrated an additional role of RNF20 beyond its previously implicated functions in transcription. Following DSB recognition by the MRN complex, activated ATM phosphorylates both RNF20-RNF40, which transiently remain at the

DSB sites to allow chromatin decompaction and recruitment of additional repair proteins (Shiloh et al., 2011). RNF20-40 mediated histone H2B monoubiquitination promotes RAD51-dependent HRR (Nakamura et al., 2011). Independently of RAD51, the RAD51 paralog, XRCC3, accumulates at DSBs and facilitates the formation of RAD51 nucleoprotein filaments (Forget et al., 2004). Since XRCC3 interacts with RNF20 based on our BC-PIN, we propose that RNF20 may play additional roles in regulating HRR. This may include promoting the binding of XRCC3-RAD51 to mediate HRR or XRCC3 may itself be a substrate of RNF20/40 (Figure 22). Monoubiquitination of XRCC3 could promote its interactions with RAD51 or stabilize its complex formation with its binding partner RAD51C. Public mass spectrometry data supports this possibility as ubiquitination of XRCC3 in its RAD51-like domain has been reported (www.phosphosite.org).

Another interesting RNF20 interaction derived from the BC-PIN is with the CHK2 kinase. Like ATM, CHK2 is activated by DNA DSBs and plays a central role in DNA damage signal transduction (Stracker et al., 2009). Activated CHK2 phosphorylates a number of distinct downstream effectors (BRCA1, p53, E2F1, Cdc25A and Cdc25C), which result in the activation of cell-cycle arrest, DNA repair, senescence or apoptosis (Antoni et al., 2007). Notably, ATM and CHK2 phosphorylate a number of common substrates, including p53, where it is clear that both phosphorylation events contribute to its full activation (Chao et al., 2006). Therefore, the association of RNF20 with CHK2 could indicate that full RNF20 activation also depends on its phosphorylation by CHK2 (Figure 22). Another intriguing possibility is that RNF20 couples the RAD51C/XRCC3 dependent activation of CHK2 by facilitating protein-protein interactions or formation of macromolecular complexes (Badie et al., 2009). Collectively, our findings suggest additional roles for RNF20 in DNA damage signaling and give insights into its possible mechanism of action at DNA damage sites.

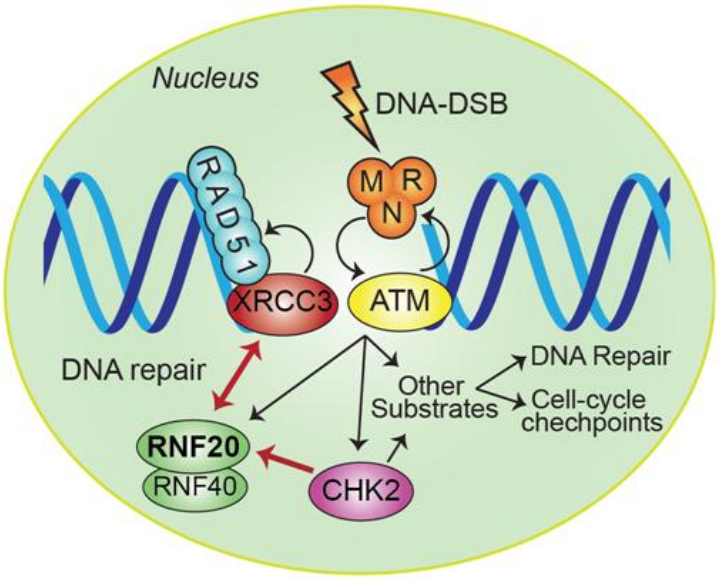


Figure 22. Model for a novel role of RNF20 in DNA repair.

RNF20 can promote RAD51-mediated HRR. Upon DNA damage, the MRN complex first activates ATM, which in turn activates the RNF20-RNF40 dimer (black arrows). Based on our two novel RNF20 interactions with XRCC3 and CHK2 (red arrows), we hypothesize that RNF20 may foster DNA repair mechanisms by favoring the XRCC3 binding to RAD51 to subsequently facilitate the formation of RAD51 nucleoprotein filaments that mediate HRR. Furthermore, the association of RNF20 with CHK2 could also act as a relay mechanism for the ATM-CHK2 signaling, ensuring the DNA damage repair.

3.3.4 FAM84B is involved in control of DNA repair mechanisms

The Breast cancer membrane protein 101 (FAM84B) was originally identified in EGF receptor positive cell lines. It has no functionally characterized homologues, representing a completely unique BC membrane protein. Although it was initially annotated as a membrane protein, cellular location studies have demonstrated widespread intracellular localizations and it appears in the plasma membrane at high levels, particularly in areas of cell-cell contacts (Adam et al., 2003). Furthermore, RT-PCR and immunohistochemical analysis of FAM84B expression have showed very low expression levels in multiple normal tissues. In contrast, high levels of mRNA and protein expression have

been detected in breast carcinoma cells relative to normal breast tissue, with more than 2-fold upregulated mRNA levels (Rhodes et al., 2007). It has been speculated to play a role in breast tumor development by blocking the tumor suppressor function of α1-catenin (Adam et al., 2003).

Interestingly, we found two interactions involving the FAM84B protein with the DNA DSB repair proteins CHK2 and RAD51, which were validated by co-IP and co-localized in the nucleus (Figure 19B-C). We also analyzed the co-expression of these genes in several tissues (see Materials and Methods), where they showed a high correlation (r^2 = 0.72 for CHK2-FAM84B; r^2 = 0.89 for RAD51-FAM84B). In addition, we found that depletion of FAM84B by siRNA in human cells significantly increased cell sensitivity to ionizing radiation (Figure 19B), providing evidence that FAM84B is indeed related to the DNA damage response.

Based on these observations, and since FAM84B is not predicted to be a phosphorylation substrate of CHK2 (according to prediction algorithms NetworKIN (Linding et al., 2007) and KinasePhos 2.0 (Wong et al., 2007), we speculate that FAM84B might affect CHK2 kinase signaling through physical interactions. Furthermore, FAM84B may stimulate DNA repair through of its association with RAD51. Notably, we were only able to detect the FAM84B-RAD51 interaction by co-IP of endogenously expressed proteins after cell irradiation, indicating that this interaction exclusively occurs upon DNA damage (Figure 23A). We did not detect the recruitment of FAM84B at DNA damage sites, meaning that the interaction with RAD51 may not be occurring at DSBs. In fact, numerous studies have shown the presence of cytoplasmic RAD51 in normal proliferative cells (Davies et al., 2001), suggesting that the nuclear levels of RAD51 are regulated by coordinated changes in its subcellular distribution, mainly by an intricate interplay between transcriptional regulation and protein turnover during the cell cycle (Gildemeister et al., 2009). We thereby investigated the subcellular distribution of RAD51 and FAM84B proteins before and after IR exposure. To this end, cells were lysed and the mitochondrial, cytosolic and nuclear fractions were separated and analyzed for the presence of FAM84B and RAD51, respectively. Interestingly, we observed a 1.48-fold increase in cytosolic FAM84B expression upon radiation (Figure 23B-D).

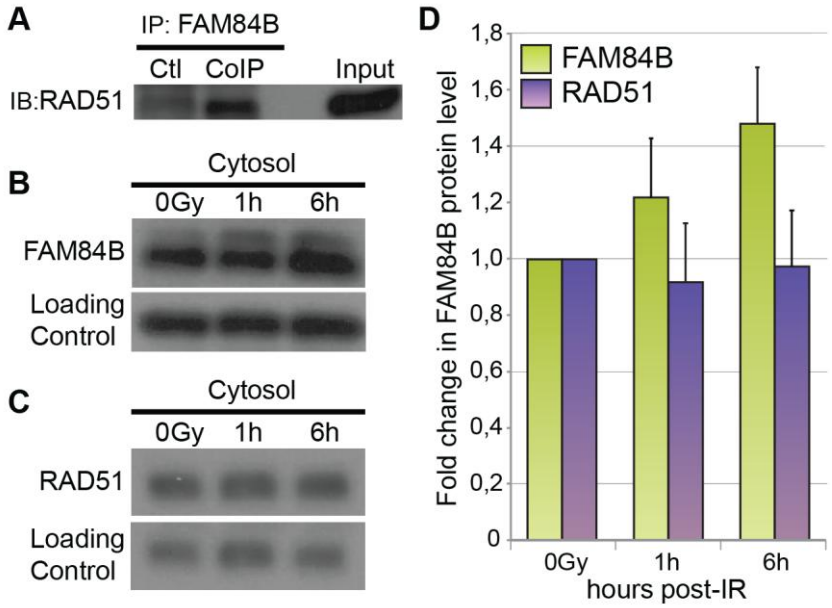


Figure 23. Upregulation of cytosolic FAM84B upon IR.

(A) FAM84B-RAD51 interaction examined by co-IP experiments upon IR-cell damage. FAM84B was immunoprecipited and RAD51 immunoblotted using specific primary antibodies. **(B)** Analysis of FAM84B expression in HEK293T cytosolic fractions 1h and 6h post-irradiation. **(C)** Conversely, RAD51 expression revealed no changes. The Western blots are representative of triplicate experiments. **(D)** Based on them, cytosolic FAM84B showed an overall 1.2- and 1.5-fold increase at one and six hours following damage, respectively. Furthermore, applying a linear expression model (expression ~ Beta • TimePoint + Gamma), we observed a positive trend with a P-value = 0.02. No trend was observed in RAD51 expression.

Conversely, the cytosolic distribution of RAD51 was not affected after DNA damage (Figure 23C-D), showing that IR generates an accumulation of FAM84B in the cytosol while RAD51 levels are unaltered, consistent with previous findings that RAD51 expression is not affected by DNA damage (Chen et al., 1997). The interaction of FAM84B with RAD51 could prevent the formation of RAD51 filaments in the cytoplasm, which could otherwise compromise their nuclear entry (Gildemeister et al., 2009). This association would thus be a mechanism to maintain the available pool of RAD51 to ultimately boost HRR. Our survival assays provide further support to these presumptions, since downregulated expression of FAM84B induces cell radiosensitivity, indicating a diminished DNA damage response.

Considering these possibilities, our findings suggest that up-regulation of FAM84B in breast cells could promote early tumorigenesis by altering DNA repair mechanisms via RAD51 stabilization. In addition, FAM84B might also have an impact in tumor progression by perturbing DNA-damage response signals associated with the CHK2 kinase pathway. Thereby, inhibition of FAM84B might prove efficacious as a therapeutic target when used in combination with DNA-damaging chemotherapeutic drugs or, alternatively, in radiotherapy as an adjuvant treatment to enhance tumor radiosensitivity. Notwithstanding, considerable work needs to be done to fully elucidate the relevance of FAM84B in DNA damage repair and its clinical efficacy against breast cancer.

CHAPTER

II

IDENTIFICATION AND VALIDATION OF NOVEL BC DRUG TARGETS AND DRUG COMBINATIONS

4. Introduction II

4.1 The Drug Discovery Process

Drug discovery is a time consuming and expensive process. Estimates of time of currently bringing a new drug to market is between 10-17 years and cost on average $1.8 billion (Paul et al., 2010). In addition, development of a new drug is a risky process, as about nine out of ten candidate molecules fail to complete the course before they are accepted as drugs (Shah and Federoff, 2009).

The drug discovery process is schematically presented in Figure 24, and it involves several steps. Previously, there must be an investigation of the biochemical, cellular and pathophysiological mechanisms behind a certain disease. The first step in the development process is the identification and validation of a molecular candidate drug target. A target is a broad term which can be applied to different biological entities such as proteins, genes and RNA. A good target needs to be efficacious, safe, meet clinical and commercial needs and, above all, 'druggable'.

Target validation is a crucial step in the drug development process. Most drugs act by inhibiting of the action of a particular target protein, but the only way to be completely certain that a protein is relevant in a given disease is to test the idea in humans. Obviously such clinical trials cannot be used for initial drug development, which means that a potential target must undergo a validation process: its role in disease must be clearly defined (Smith, 2003). Validation techniques range from in vitro tools through the use of whole animal models and, while each approach is valid, confidence in the observed outcome is increased by a multi-validation approach.

As soon as a target has been validated to have a potential impact for a special disease, the search for compounds that interact with the target starts. During the hit identification and lead discovery phase the compound screening assays

are developed. A 'hit' molecule can be defined as a compound which has the desired activity in a compound screen and whose activity is confirmed upon retesting (Hughes et al., 2011). The medicinal chemistry part, screening, hit to lead and lead optimization, aims to identify and optimize compounds towards a potential drug.

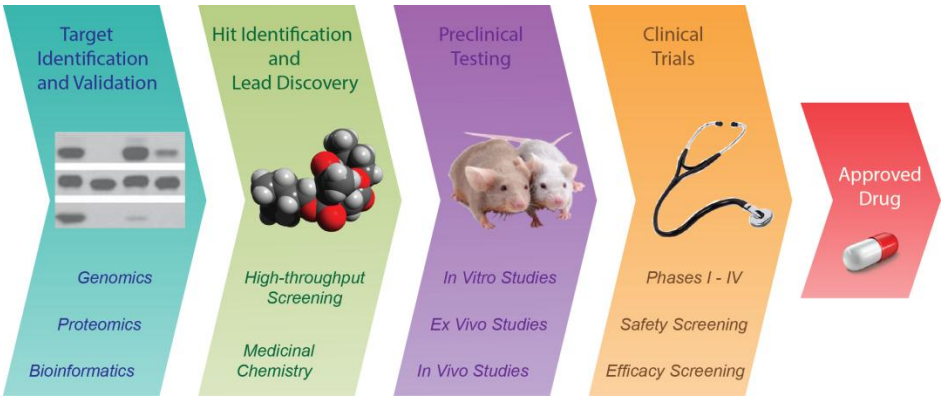


Figure 24. The drug discovery process.
The different phases of modern drug discovery and development.

Once a compound that fulfills all of these requirements has been identified, it will begin the process of drug development prior to clinical trials. In the preclinical phase, both in vitro and in vivo tests are performed to determine drug safety, dosing and toxicity before it goes to human testing. After animal testing, the molecule is tested on human volunteers (clinical phase I/II/III) and, if the drug shows a beneficial effect in relevant patient groups without major side effects, the drug can then get the final approval and be a marketed medicine.

4.2 Breast Cancer Current Therapy

Different types of treatment are available for patients with breast cancer (Figure 25), depending on tumor and patient characteristics. After diagnosis, most breast cancers are primarily treated with surgery to remove the tumor alone (lumpectomy) or the entire breast (masectomy). Surgery is followed by

radiotherapy, a treatment that uses high-energy radiation to kill cancer cells or to prevent them from growing further.

After surgery and/or radiotherapy, breast cancer treatment is usually continued by chemotherapy and/or targeted therapy, depending on subtype and progression of the disease. Chemotherapy is a systemic treatment that uses cytotoxic drugs to kill or stop growing of cancer cells, and that can be either used before (neoadjuvant) or after (adjuvant) surgery (primary treatment). Chemotherapeutic agents act by killing indiscriminately cells that divide rapidly; thus, impairing cancer cells but also against healthy cells of bone narrow, hair follicles or digestive tract, resulting in characteristic and strong side effects (immunosuppression, inflammation of the digestive tract and alopecia).

As breast epithelial cells proliferate under hormonal control, breast cancer ER positive tumors are often treated by adjuvant hormonal therapy (with Tamoxifen or an aromatase inhibitor). Tamoxifen is an antagonist of ER in breast tissue, and is a general regime often given to premenopausal women (Baum et al., 1983). On the other hand, aromatase inhibitors decrease the body's estrogen by blocking the aromatase enzyme from turning androgen into estrogen, and they are the first-line therapy for postmenopausal women (Howell et al., 2005).

Finally, targeted therapy refers to a new generation of cancer treatment that uses drugs designed to interfere with a specific molecular target or process that is believed to have a critical role in tumor growth or progression. This approach contrasts with traditional chemotherapy as it does not act by simply interfering with all rapidly dividing cells. For this reason, targeted therapies are expected to be more specific than current treatments and less harmful to normal cells (Sawyers, 2004). There are two types of drugs used for targeted therapy: antibody drugs and small-molecule drugs. These compounds are directed at grow factor receptors (HER1, HER2, HER3, IGFR), intracellular signaling pathways (Pi3K, AKT, mTOR, ERK), angiogenesis as well as DNA repair factors (Higgins and Baselga, 2011). Trastuzumab, for example, is a monoclonal antibody that blocks the effects of the growth factor protein HER2, therefore about one-fourth of patients (HER2+ subtype) of breast cancer may be treated with trastuzumab combined with chemotherapy (Vogel et al., 2002).

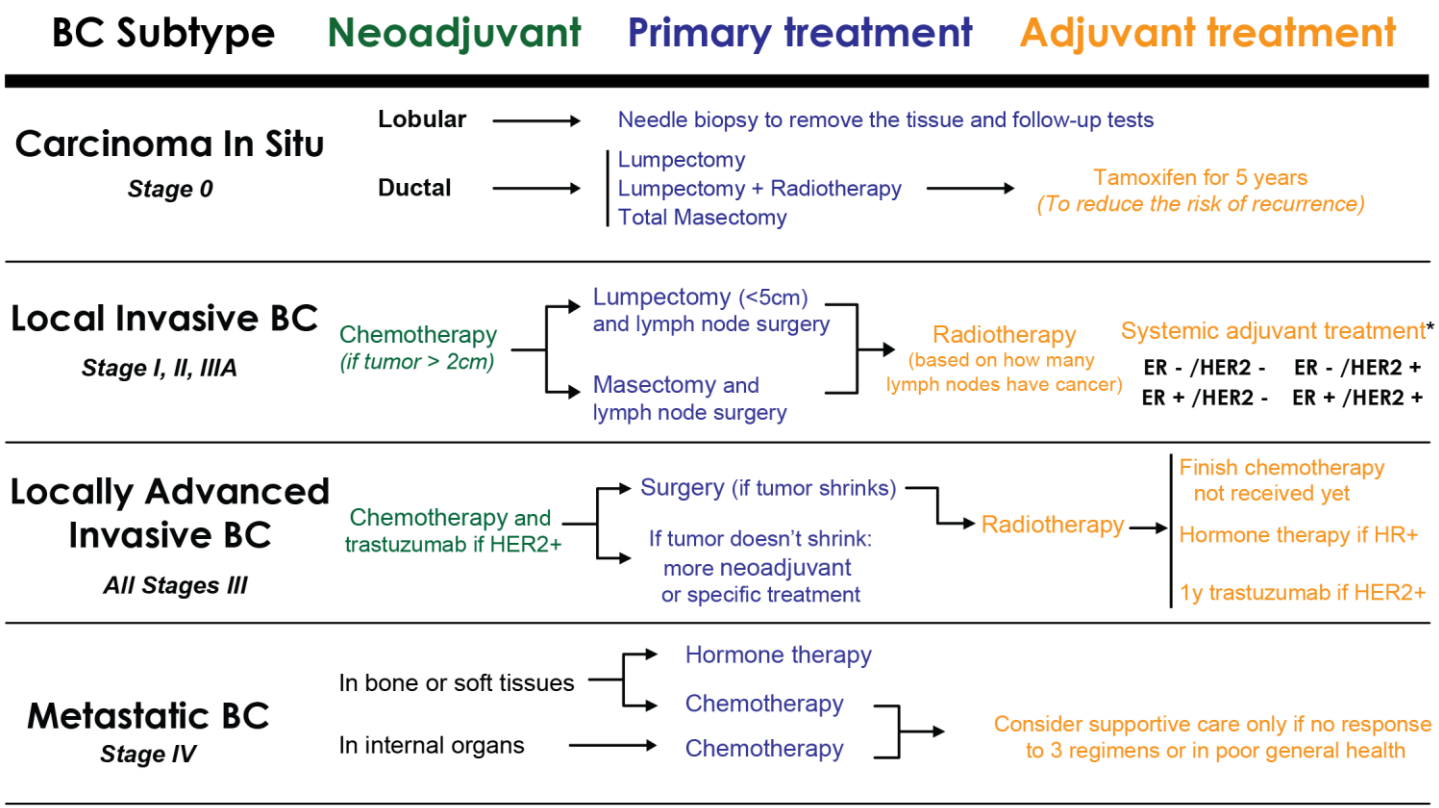


Figure 25. Main treatments for breast cancer.

Schema showing the most common strategies recommended by NCCN Guidelines™ to treat different stages of breast cancer.

4.3 Treatment Failure due to Drug Resistance

The progress accomplished in the discovery of novel anti-cancer agents has significantly improved the response rates in cancer while reducing side effects to some extends. However, it did not yet result in cure of the majority of patients, specifically, with metastatic disease. Besides, a major issue in the management of, for instance, breast cancer is the treatment failure due to drug resistance (Tsang and Finn, 2012).

Resistance to anticancer drugs may result from two main groups of factors. The first group includes host pharmacological and physiological factors such as drug absorption, metabolism and excretion, inadequate access of the drug to the tumor or inadequate infusion rate delivery (Garattini, 2007). The second group includes specific genetic or epigenetic alterations in the cancer cells. On this level, we can distinguish among primary (pre-existent) or acquired (induced by drugs) cell drug resistance (Osborne and Schiff, 2011).

While host factors are extremely important in clinical practice, we concentrate our interest in cancer cells resistance, which is fundamental in drug development. We can distinguish two types of resistance: primary and acquired resistance. On the one hand, primary resistance may be induced by mutations in the pharmacological target(s) and a lack of target dependency. On the other hand, patients with initial response to the treatment will eventually lose clinical benefits and relapse within one year due to acquired resistance (Vogel et al., 2002). Several models leading to acquired resistance have been described: loss of target expression due to continuous therapy, activating mutations downstream of the target, modifications of proteins crucial for target regulation, activation of additional mechanisms to promote cell proliferation as well as compensatory mechanisms that bypass the function of the protein which is targeted. These molecular mechanisms are known to compensate for therapeutic effects thus reducing their efficacies or leading eventually to resistance (Tsang and Finn, 2012).

In particular, signaling through alternative proteins and pathways has received considerable attention; several studies have demonstrated that cancer cells are able to adapt signaling pathway circuits upon chronic treatment by establishing alternative signaling routes through crosstalk (Bernards, 2012; Osborne et al.,

2005; Yamaguchi et al., 2013a). When considering for instance hormone therapy about 30% of the tumors show resistance towards tamoxifen treatment either due to primary or secondary resistance. Cross-talk between ER and growth factors signaling pathways, mainly the receptor tyrosine kinase (RTK)/PI3K/AKT/mTOR axis, is thought to be a mechanism contributing to endocrine resistance (Miller, 2013). Furthermore, this crosstalk is bidirectional as both estrogen and tamoxifen are able to induce HER2 phosphorylation in cells with endogenous HER2 overexpression, thereby contributing to trastuzumab resistance (Shou et al., 2004).

In conclusion, each cancer cell has a different genetic pattern resulting from the 'mutator' phenotype of most cancers (see Section 1.2.2) and, as a result, every cancer expresses a different array of drug-resistance genes (Gottesman, 2002). This heterogeneity of potential resistance mechanisms could explain why well-designed drugs developed against well-validated cancer targets nevertheless fail to deliver sustained benefit in the clinic. Therefore, treatment strategies should not be based only on the main aberrations with pathogenic importance, but also on the extensive heterogeneity present in a given tumor. Furthermore, multi-targeted therapies are needed to address resistance problem, either by using a single agent that modulate several targets (multi-targeted drugs) or by combining several agents (combinational therapy).

4.4 Combinational Therapy

Most multi-targeted therapies are developed as a mixture of agents and combinational therapy is one of the most promising strategies to prevent or delay drug resistance in cancer. As we discussed in the previous section, a key aspect in improving cancer treatment is not only to inhibit primary oncogenic signaling pathways, but at the same time to prevent the occurrence of resistance mechanisms. Hence, second-generation agents can be used to prevent the cause of these resistance mechanisms.

A combination is pharmacodynamically synergistic, additive or antagonistic if its effect is greater than, equal to, or less than the summed effects of the partner drugs (Chou, 2006). Accumulated empirical clinical experience, supported by animal models, showed that cytotoxic drugs are most effective when given in combination to achieve additive or synergistic effects (DeVita et al., 1975). More recently, drug discovery efforts have focused on combining different targeted agents, alone or with chemotherapy. Most new targeted drugs have modest efficacy with responses that are not long-lasting, thus, targeted drugs might benefit from being combined with cytotoxic agents. For example, the combination of HER2 monoclonal antibodies with paclitaxel (Slamon et al., 2001) or anti-estrogen agents with tyrosine kinase or signal transduction inhibitors (Fedele et al., 2012), have demonstrated a significant improvement in progression free survival compared to single therapy alone. Another advantage of targeted drug combinations is that they may also overcome the side effects associated with high doses of single drugs. Optimal combinational therapies could increase cancer cell killing while minimize overlapping toxicity, allowing reduced dosage of each compound (Ramaswamy, 2007).

Although combinatorial therapies are becoming the standard care in (breast) cancer treatment, the majority of drug combinations in use or in clinical trials has been discovered by clinical experience and has not been designed as such in first place. This implies that the molecular mechanisms underlying these combinations are often not elucidated, which makes it difficult to propose new combinations. Thus, identifying novel combinations with improved therapeutic effect remains a challenging task given the exhaustive number of possible combinations.

4.5 Network Biology and Cancer Therapy

An area where network biology is particularly helpful is cancer diagnosis and treatment. As discussed in Section 1.1, cancer results from a combination of multiple molecular events, thus the gene-centric drug discovery paradigm is shifting towards a network/pathway-centric approach. Despite several success stories, the adopted reductionism approach also had striking consequences. For instance, many promising drug candidates failed during the last, and most expensive, clinical phases because the action mechanisms of the pathways they target remain largely unknown (van der Greef and McBurney, 2005).

The recent development of several high-throughput technologies has provided biologists with an extensive list of genes involved in each cancer type and subtype (see Section 1.1.1 for BC). However, it does not seem enough to produce effective therapeutic strategies. Thus, the emerging challenge over the next decade is to assemble these components into functional networks and then to use these networks to understand how cancer is driven. Once we are able to understand how networks are deregulated in cancer cells, then we will able to predict how these networks might respond to drugs.

One of the major issues associated with identifying effective new drugs is the discovery of a relevant drug target (Hood and Perlmutter, 2004). However, a detailed interaction map of a given pathology may suggest potential points for therapeutic intervention (i.e. drug targets). Given the robustness of biological systems, the selection process of new putative drug targets should also consider their position in the network, preferring those nodes which are essential in the network traffic, and which are able to avoid back-up circuits that might neutralize a desired drug effect. Topological analyses of disease-associated networks and more sophisticated models deliver key information related to the pathways that are activated upon drug treatment in different cell types, providing therefore a very valuable guidance in the identification of potential points of intervention (Lee et al., 2012; Yeh et al., 2012).

Another major issue in cancer treatment is cancer resistance and, as previously discussed, the use of drug combinations to prevent or delay tumor resistance is a well-established principle of cancer therapy. A large number of molecular targeted anticancer drugs are being studied today, and it is crucial to establish

strategies to choose from the almost limitless possibilities and to focus on prioritizing the most promising combinations. Again, protein interaction networks are a perfect framework to study potential drugs that can be prescribed in combination to achieve synergistic effects. Hundreds of drug combinations have been reported, and the detailed analysis of some of them indicate that synergistic drug combinations either target the same or different proteins from the same or related pathways as well as from cross-talking routes (Figure 26) (Jia et al., 2009). Network analysis provides an understanding of disease-related pathways that is required to reveal the system response to a drug combination at the level of the cell or tissue.

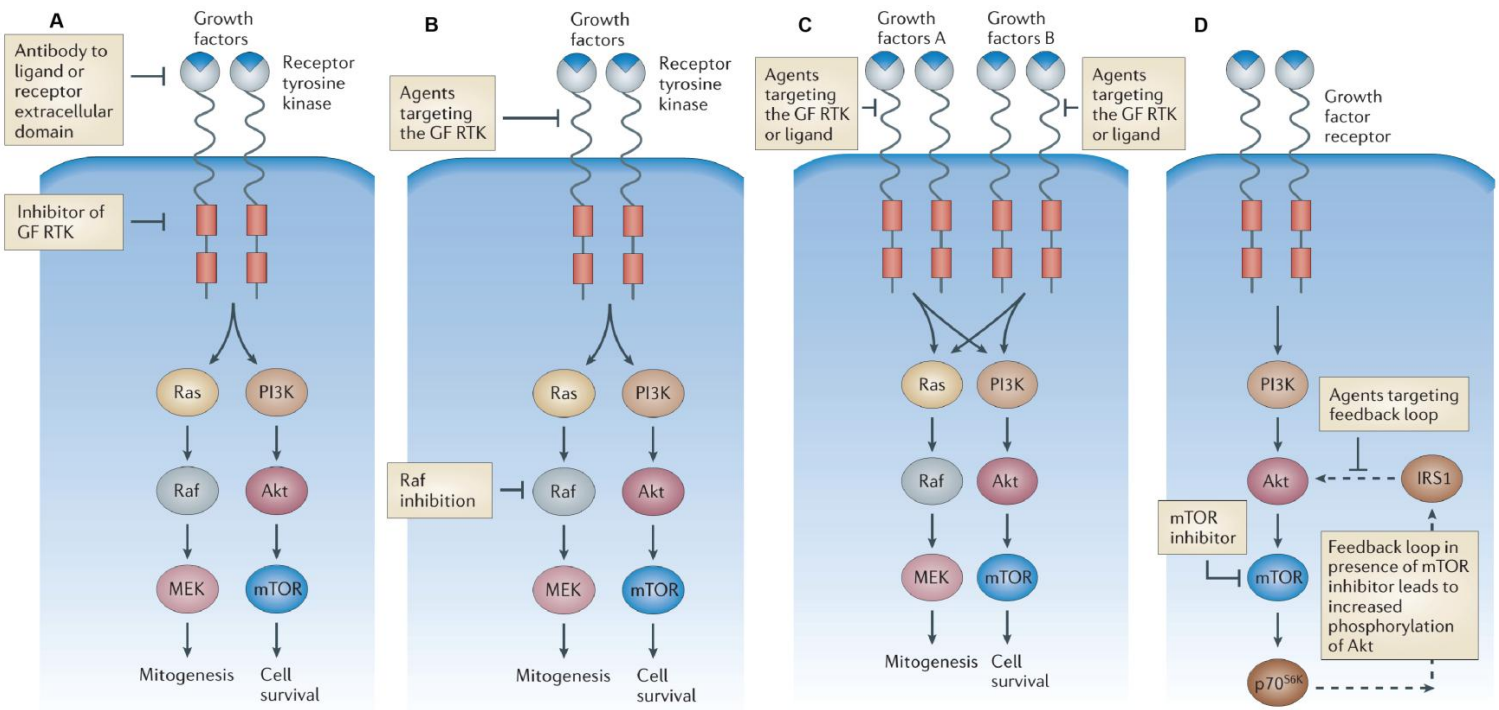


Figure 26. Strategies for optimizing combinations

(A) Maximizing the inhibition of a target such as a growth factor receptor by inhibiting both receptor-ligand binding and tyrosine kinase activity. **(B)** Maximizing the inhibition of a pathway by blocking a series of signaling components within the pathway. **(C)** Inhibition of parallel pathways by blocking two growth factor receptors or inhibiting downstream components in parallel pathways. **(D)** Inhibition of a target and the feedback loop that result in resistance. *From Dancey and Chen, 2006*.

5. Objectives II

The main objectives of this chapter of the thesis are:

- To identify novel putative drug targets for breast cancer using network topology properties and to perform an *in vitro* assessment of the effect from inhibiting these targets in BC.
- To validate *in vitro* a set of combinations which involve novel putative drug targets with current breast cancer treatment drugs, and that have been predicted based on a pathway crosstalk inhibition measure.
- To predict and validate novel combinations involving two current breast cancer treatment drugs using the same approach.

6. Results and Discussion II

6.1 A network biology approach towards novel BC drug targets identification

Conventional cytotoxic chemotherapeutic drugs, including microtubule modulators, DNA interactive agents and anti-metabolites, block tumor cells by impairing processes or molecules necessary for replication. Target-oriented drugs, on the other hand, inhibit cancer growth by interfering with molecules explicitly implicated in tumor growth and progression, such as proteins involved in cell invasion, metastasis, apoptosis and tumor-induced angiogenesis. In this work, I initially validated a set of novel therapeutic targets that were identified by applying a network biology approach.

Several studies indicate that the analysis of local and global network characteristics of known targets is a promising tool for assessing a protein's drug status (Emig et al.; Zhu et al., 2009). The cellular context of therapeutic and candidate targets in terms of their underlying pathways and networks was considered in the approach. The main idea of our strategy was to exploit drug targets of breast cancer agents whose functional annotation reflect frequently perturbed mechanisms that need to be modulated effectively.

Starting from previously compiled list of 59 breast cancer driver genes (see Section 3.1.1), we first generated a breast cancer specific interaction network by integrating gene products which interact directly or indirectly with these proteins. We then identified proteins of the network that are targeted by current breast cancer therapeutics, and we compiled an exhaustive set of candidate proteins which are involved in similar biological processes and pathways as these current breast cancer drug targets. We determined the functional similarity between a protein and a breast cancer drug, according to its targets, by employing a combined similarity measure using pathway and GO annotations (biological process). Hence, 210 proteins with statistically

significant functional similarity to at least one breast cancer drug (P-value < 0.0015) were considered as candidate targets. Next, we ranked the identified candidate targets according to their topological network similarity to breast cancer targets to distinguish candidates resembling not only the functional but also the topological characteristics of breast cancer targets. The complete workflow of the drug target identification strategy is shown in Figure 27.

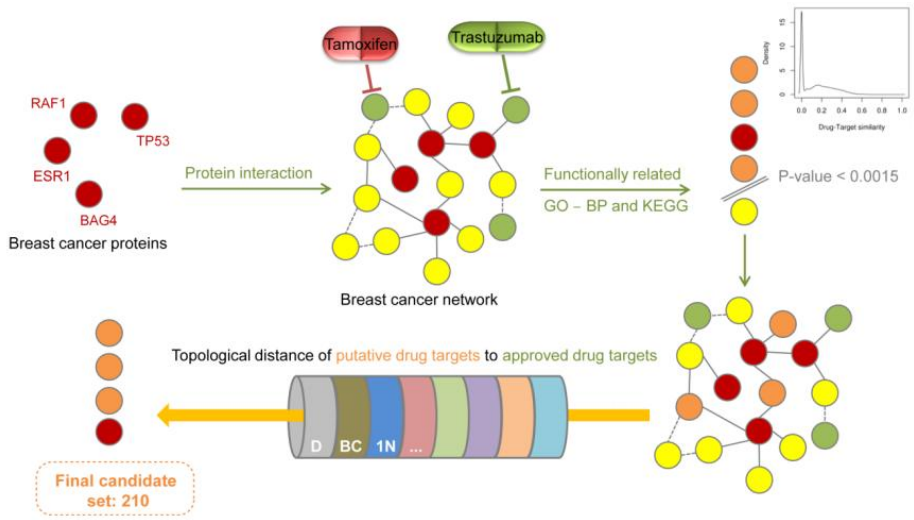


Figure 27. Overview on the workflow of the network biology approach for identifying novel breast cancer therapeutic targets.

Our method comprises three main steps: (1) generation of a breast cancer specific interaction network, (2) identification of candidate proteins involved in similar biological processes and pathways as current breast cancer targets, and (3) prioritization of the candidate targets based on their topological network similarity to breast cancer drug targets.

The 210 putative drug targets were further analyzed to extract the most promising targets for tackling breast cancer efficiently. Firstly, we excluded nine of the research targets, e.g., MAP3K4, ERBB3, DNMTB3 and NRG1 because they are currently under investigation for breast cancer treatment. We also examined whether, based in literature evidence, these targets might be implicated in the development or progression of breast cancer, and we excluded the proteins without clear functional evidence and a low likelihood of being essential in breast cancer. We generated a final list of 100 putative promising novel BC targets. This part was done by Dr. Samira Jaeger in the lab and will be described in Materials and Methods chapter.

6.2 *In vitro* assessment of target inhibition

The fundamental aim of our target identification strategy is the validation of predicted candidates and combinations, i.e., demonstrating the effect of modulating protein activities individually and in combination through experiments. In order to avoid problems related to the combination of siRNA/shRNA transfection and drug treatment on cell lines, we decided to focus on the targets with an associated chemical modulator. After considering the availability of these chemical modulators the final list of candidate targets was reduced from 100 to 54.

NCI 60 validation

Before performing a traditional experimental validation, we checked whether our candidate targets have been assessed in one human cell line panel. Resources, such as the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), the NCI60 (Shoemaker, 2006) or other publicly available large-scale cell line collections (Garnett et al., 2012; Greshock et al., 2010; Neve et al., 2006), are commonly used to profile cancer cell lines for various genetic abnormalities and gene expression changes at the DNA, RNA and chromosomal level as well as for testing their sensitivity towards approved/experimental drugs and other chemical molecules. We selected the NCI60 repository as one of the most comprehensive sources of compound activity profiles in human cancer cell lines. The NCI60 database contains currently 49,282 publicly available compound activity profiles for 59 human cancer cell lines (including 7 BC cell lines).

Our objective was to evaluate the effect observed in the chemicals known to modulate our 54 novel putative targets in the human BC cell lines. To this end, we first mapped these ligands to NCI compounds based on their CAS numbers, if available, or by computing chemical similarity using their 2D structures (see Materials and Methods). Considering NCI compounds with identical CAS numbers or with a chemical similarity of 1.0, we mapped 109 ligands to 102 NCI compounds covering 33 candidate targets. Next, we analyzed the activity profiles of these NCI compounds, by obtaining three parameters: (1) GI50: 50% inhibition of the cell growth (2) TGI: total growth inhibition, and (3) LC50: 50% of the cancer cells killing.

In order to assess the significance of the candidate target activities obtained we compared them with activities measured for current approved and experimental breast cancer drugs ($NCI_{BCdrugs}$), for the complete set of NCI compounds ($NCI_{Background}$) and for a set of NCI compounds derived from proteins randomly selected from the breast cancer network ($BC_{Network}$).

In Figure 28, we show the comparison of the activity data among the four different data sets. From this analysis, we can conclude that compounds mapped to ligands which inhibit predicted candidate targets (1) yielded significantly better responses in BC cell lines than random molecules from NCI60, (2) induced significantly better responses in most conditions than ligands targeting randomly selected proteins from the BC network (as expected this difference is less pronounced as detected for the NCI background), and (3) showed similar responses as observed for approved and experimental breast cancer drugs. These results emphasize the potential relevance of the predicted candidate drug targets with respect to inhibit cell growth and to kill breast cancer cells.

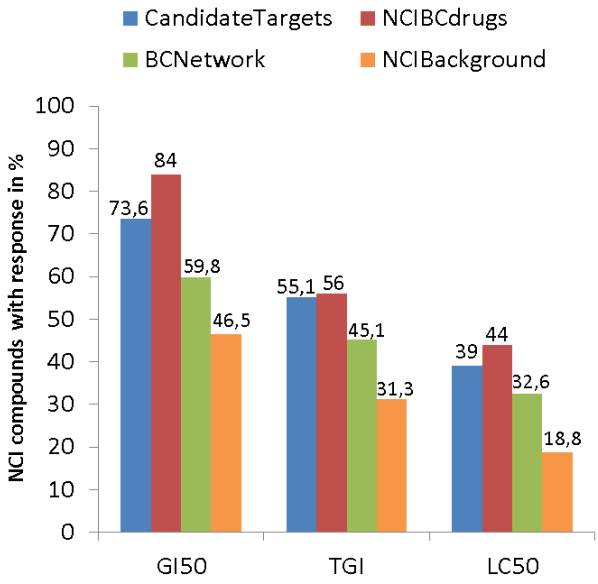


Figure 28. NCI60 screening results

Comparison of the cytostatic/cytotoxic activities derived for candidate targets, breast cancer drugs, the NCI as background and randomly selected proteins from the breast cancer specific network.

MTT validation

Finally, we experimentally tested by MTT assays (see Materials and Methods) a selected set of candidate drug targets. The objective was to assess whether the selected candidate targets do have an impact on cell growth in breast cancer. We exploited four different types of information for candidate targets prioritization; (1) literature evidence, (2) the number of driver pathways that are predicted to be affected after inhibiting each target (driver pathway coverage), (3) topological similarity to known breast cancer targets (by means of their rank) and (4) activity profiles retrieved from the NCI60 cell line.

Next, we selected the best associated compounds to inhibit each target. As our work is target-centered, we carefully selected chemical compounds with high specificity and binding affinity. Table 5 shows the final list of eight candidate targets selected for experimental validation as well as each associated compound used.

Target	Compound	CAS Number
IL1R1	Anakinra (Kineret)	143090-92-0
MAP2K2 (MEK2)	SL 327	305350-87-2
MAP3K7 (TAK1)	5Z-7-Oxozeaenol	66018-38-0
MAPK7 (ERK5)	XMD8-92	1234480-50-2
PIK3CB	TGX-221	663619-89-4
PPP2R5A	Norcantharidin	29745-04-8
PTPN6 (SHP-1)	Sodium stibogluconate	16037-91-5
RAP1A	GGTI-298	1217457-86-7

Table 5. Candidate drug targets selected for experimental validation.

Each target was tested in six different cell lines (Table 6). We chose four different breast cancer cell lines that represent the major subtypes of breast cancer (hormone receptor positive, HER2 overexpressed, triple positive and triple negative breast cancers). In addition to the receptor status, these cell lines

harbor different genetic aberrations that are specific for the main breast cancer subtypes. We also wanted to evaluate whether our targets could be also relevant to study another type of cancer, to this end we performed the assay using an osteosarcoma cell line (U2OS). Finally, we tested a non-tumorigenic cell line (MCF-10A) that served us as a negative control.

Cell line	Features				Histology Subtype
	Subtype	ER / PR	HER2 (ERBB2)	Other Mutations	
MCF-7	Luminal	+	0	CDKN2A / PIK3CA	Metastatic adenocarcinoma
SK-BR-3	Luminal	0	+	TP53	Metastatic adenocarcinoma
BT474	Luminal	+	+	PIK3CA / TP53	Ductal carcinoma
MDA-MB-231	Basal	0	0	BRAF/ CDKN2A / KRAS / NF2 / TP53	Metastatic adenocarcinoma
U2OS	Osteo-sarcoma	IGF1R and IGF2R expressed		-	Tibia sarcoma
MCF-10A	Non-tumorigenic epithelial cell line			-	Mammary gland

Table 6. Features of cell lines used in candidate target validation.

Given the data from the MTT assays we model the respective dose-response curves using sigmoidal fitting to from which we determined the half-maximal inhibitory concentration (IC_{50}) for each target and cell line (see Materials and Methods). Here, the IC_{50} represents the drug concentration causing 50% inhibition of the cell population compared to non-treated cells.

Note that the obtained IC_X (where X represents the percentage inhibitory effect) values are always associated with a standard error. In cases in which the determined IC_X lies outside of the tested range of drug concentrations, we

exploit this standard error to decide whether we consider this IC_X to be valid. As not all drugs were effective, we were unable to determine all IC_{50} values with an error lower than 0.15. Table 7 shows these IC_{50} whether they can be determined as well as the maximum inhibition (IC_{MAX}) that can be reached by using this single target. In summary, the results we obtained for each target ordered from lower to higher effect were:

- IL1R1 and PTPN6 inhibition did not lead to cell growth inhibition in none of the cell lines.
- MAP2K2 inhibition yielded an 80% reduction of cell growth but only in MCF-10A control cell line and with a very high dose (76.5 μM). Besides, we reached IC_{40} = 40.4 μM in triple-negative (MDA-MB-231) and we observed a low effect (IC_{20}) in other two BC cell lines (MCF-7 and BT474).
- PPP2R5A inhibition had no effect in MCF-7 and SK-BR-3 cell lines, while it reached IC_{50} and $IC_{60/70/80}$ as maximum cell growth inhibition in four out of six cell lines, but the drug doses required for this effect are extremely high. The most remarkable effect was detected in the osteosarcoma cell line (U20S), which showed much lower IC_{50} (36 μM) and IC_{60} (64.7 μM) doses.
- PIK3CB inhibition yielded around 40-60% inhibition in all cell lines. Take to account that for this target we observed a higher experimental variability, especially in MCF-7, BT474, U2OS and MCF-10A (IC_X values are shown with the * character in Table 7).
- MAPK7 inhibition yielded an IC_{50} and IC_{80} in all cell lines except for BT474 (triple positive). The strongest effect was observed in hormone positive cell line (MCF-7), with IC_{50} and IC_{80} values of 6.0 and 28.5 μM, respectively.
- RAP1A inhibition yielded an effect in all cancer cell lines but no effect was measured in MCF-10A control cells. In BT474 cells the maximum inhibition was 40%, but in the remaining four cell lines these inhibition was much stronger, with IC_{50} values ranging from 10.4 to 18 μM, and reaching IC_{80} in all four cases.
- MAP3K7 was the only target whose inhibition was able to consistently reduce cell growth by 50% and 80% in all the cell lines.

Therefore we can distinguish three groups of genes: three genes that had zero or negligible effect (IL1R1, PTPN6 and MAP2K2), two more genes with an intermediate effect (PPP2R5A and PIK3CB) and three genes whose effect was extremely high in general (MAPK7, RAP1A and MAP3K7).

We centered our attention in the last three targets. In the three cases MCF-7 was the most sensitive cell line (with lower IC_{50} and or IC_{80}), but only MAP3K7 inhibition was able to reach IC_{50} in BT474. MAPK7 is a member of the MAP kinase family that acts as an integration point for multiple biochemical signals, and it is involved in cellular processes such as proliferation, differentiation and cell survival (Kato et al., 1998). RAP1A belongs to the family of Ras-related proteinsand plays an important role in adhesion and migration of lymphocytes (Beraud-Dufour et al., 2007). And MAP3K7 kinase mediates the signaling transduction induced by TGF beta and morphogenetic protein (BMP), and controls a variety of cell functions including transcription regulation and apoptosis (Wang et al., 2001). Furthermore, we found that two of these targets have been already tested in cancer, however, our approach studies specifically the role of these targets in each subtype of breast cancer.. Yang et al., 2010 demonstrated that MAPK7 inhibition blocked tumor cell proliferation *in vitro* and significantly inhibited tumor growth *in vivo* by 95% using a murine lung cancer model. RAP1A has been also recently tested in lung cancer cell lines: Du et al., 2012 demonstrated RAP1A targeting sensitizes NCI-H1155 cells to paclitaxel. Conversely, MAP3K7 has never been tested in a proliferation assay of cancer cells.

When analyzing the different behaviour among the cell lines we noticed that the triple positive BT474 was the most resistent cell line to the compounds tested, as only MAP3K7 inhibition was able to reach IC_{50} using a low compound dose. Similarly, only the three best compounds (MAPK7, RAP1A and MAP3K7 inhibitors) were able to reach IC_{50} in the hormone positive MCF-7. On the other side, the triple negative MDA-MB-231 and the control MCF-10A cell lines were very sensitive in general to the targets inhibition, although the relevance of the results obtained in MCF-10A cell line are controversial (see section 6.4).

Target	MCF-7 (ER/PR+)		SK-BR-3 (HER2+)		BT474 (Triple +)		MDA-MB-231 (Triple -)		U2OS (Osteosarcoma)		MCF-10A (non-tumorigenic)	
	IC_{50}	IC_{MAX}	IC_{50}	IC_{MAX}	IC_{50}	IC_{MAX}	IC_{50}	IC_{MAX}	IC_{50}	IC_{MAX}	IC_{50}	IC_{MAX}
IL1R1												
MAP2K2		IC_{20} = 4.82				IC_{20} = 27.03		IC_{40} = 40.37			21.70	IC_{80} = 76.48
MAP3K7	2.88	IC_{80} = 9.46	7.00	IC_{80} = 20.42	6.70	IC_{80} = 20.48	2.14	IC_{80} = 14.53	4.53	IC_{80} = 10.54	5.22	IC_{80} = 10.91
MAPK7	6.02	IC_{80} = 28.46	24.1	IC_{60} = 35.09			15.68	IC_{80} = 42.18	23.90	IC_{60} = 35.27	12.24	IC_{80} = 26.42
PIK3CB		IC_{40} = 13.62*	36.30	IC_{60} = 51.98	35.62*		24.09			IC_{40} = 34.41*	44.26*	
PPP2R5A					95.8	IC_{60} = 110.45	91.17	IC_{70} = 110.55	35.98	IC_{80} = 64.70	58.54	IC_{80} = 119.96
PTPN6												
RAP1A	10.4	IC_{80} = 15.49	18	IC_{80} = 30.60		IC_{40} = 21.53	17.60	IC_{80} = 21.20	15.25	IC_{80} = 18.61		

Table 7. Summary of effects obtained in the candidate target validation.

If determined, IC_{50} values are shown. The maximal inhibitory effects are also reported, and they are depicted as grey when the maximum effect was lower than 50%. Units = μM.

* PIK3CB inhibitory assays showed an important experimental variability in MCF-7, BT474, U2OS and MCF-10A cell lines.

6.3 A network biology approach to predict novel drug-drug and target-drug combinations

There is growing evidence that crosstalk between alternative signaling pathways, is one of the mechanisms for resistance to therapy in cancer (Osborne et al., 2005; Yamaguchi et al., 2013b). Combinatorial therapies which target distinct adaptive signaling mechanisms simultaneously are evolving as key strategies for improving the management of breast cancer with respect to resistance. Given the role of alternative signaling, i.e., pathway crosstalk, in drug resistance and therapy failure, we proposed a strategy for inferring target-drug and drug-drug combinations tackling particularly this problem. For this reason, Dr Samira Jaeger developed a computational approach for inferring drug combinations regarding drug resistance by taking functional redundancy and pathway crosstalk into account. Our strategy was based on a crosstalk inhibition measure, which is able to quantify the amount of pathway crosstalk that can be prevented by inhibiting two (sets of) targets in combination. My job was to experimentally validate a subset of 10 drug-drug combinations (DC01-DC10) and 13 novel target-drug combinations (TDC01-TDC13) predicted as potentially synergistic and/or with a significant crosstalk inhibition.

Drug-drug combinations

We first collected a list of 64 agents that were either approved (U.S. National Cancer Institute, NCI) or experimentally studied (Therapeutic Target Database, TTD (Zhu et al., 2012) for breast cancer treatment. We generated all possible pair-wise combinations from these 64 agents and studied which of them are either in use or tested in clinical trials. Of the 2,016 combinations, 166 are documented by the ClinicalTrial.gov, the U.S. Food and Drug Administration (FDA) orange book, the NCI or the Drug Combination Database (DCDB) (Liu et al., 2010), and thus are considered to be efficient combinations. The remaining combinations constituted a set of random combinations. Given the two sets, we computed the pathway crosstalk inhibition between the involved drugs as described in Materials and Methods, and we observed that combinations in use have a significantly higher impact on pathway crosstalk than random combinations, with a Crosstalk Inhibition (CI_C) average of 0.34

compared to 0.25 (P-value = 8.96x10-6). This implies that the proposed crosstalk inhibition measure provides a valuable tool for identifying drug combinations.

We exploited the potential pathway crosstalk inhibition for drug combinations from the random set. Figure 29 shows the crosstalk inhibition among all pairs of breast cancer drugs including the ones currently used (specifically highlighted). Overall, we predicted 384 novel drug combinations exceeding the crosstalk inhibition threshold of 0.34. In addition, 35 combinations, from

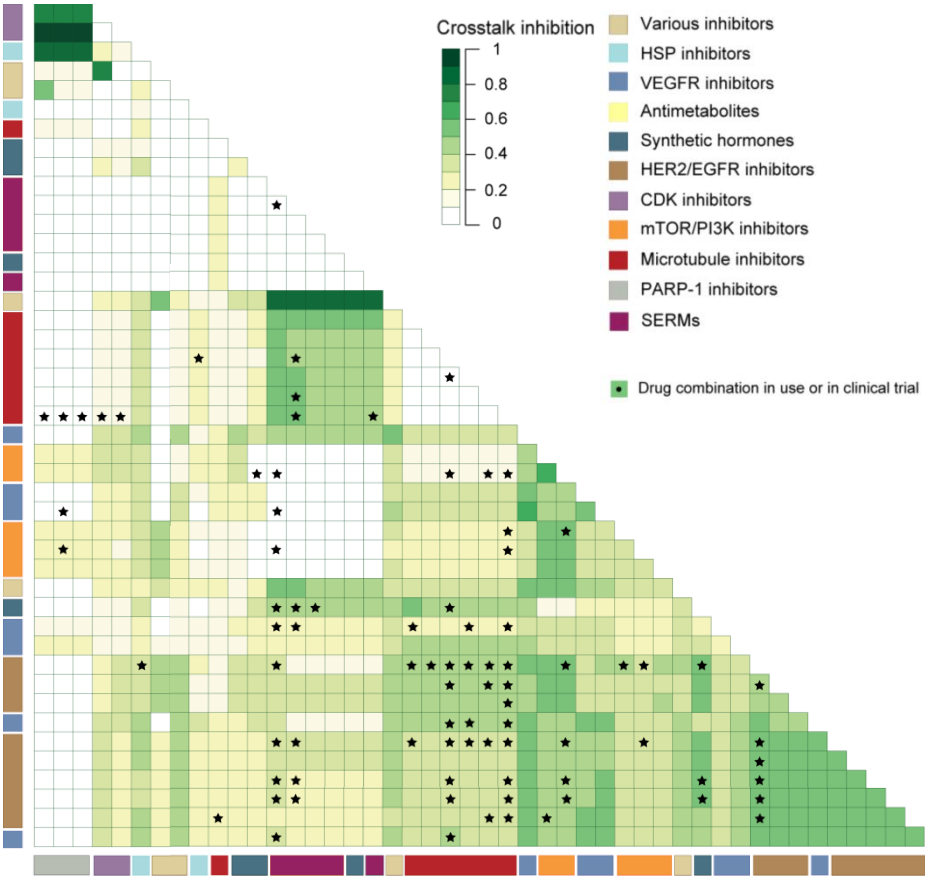


Figure 29. Crosstalk inhibition computed between BC drugs.

Crosstalk inhibition determined between breast cancer therapeutics, both approved and experimental. Drug combinations in use are indicated with a black star symbol (★).

which 29 did not reach the inhibition threshold, showed a synergistic/additive behavior in the crosstalk inhibition, that is, the combinatorial inhibition being larger than the sum of the inhibition determined each single drug: $CI_C \geq CI_{D1} + CI_{D2}$. Including these potentially synergistic combinations, the predicted set of drug-drug combinations comprised 413 novel drug combinations.

Target-drug combinations

In a similar manner as described for the drug-drug combinations, we simulated the impact of inhibiting novel candidate targets previously predicted in Section 6.1 in combination with breast cancer drugs. We again exploited the potential pathway crosstalk inhibition for drug combinations, and two outcomes were desirable: (1) Increased efficacy of the drug due to the inhibition of oncogenic pathways at different points, and (2) prevention of alternative signaling routes that might confer to drug resistance. Figure 30 shows the crosstalk inhibition determined for candidates in combination with the set of breast cancer drugs. Overall, 74 candidates were potentially beneficial for diminishing pathway crosstalk with at least one of the breast cancer drugs. In total, 2,367 combinations had an impact on pathway crosstalk, and 1,108 of these combinations exceed the CI_C threshold of drug combinations in use (0.34).

As we were interested in candidates with high individual or complementary contribution to the crosstalk inhibition, we sought for candidates with a high benefit of the candidate target (BCT). For example, a BCT $\geq$ 1.5 corresponds to a 1.5-fold increase of the overall inhibition by additionally inhibiting the candidate target. This applies for 421 combinations of which 137 have a crosstalk inhibition > 0.34, and thus present promising target-drug combinations.

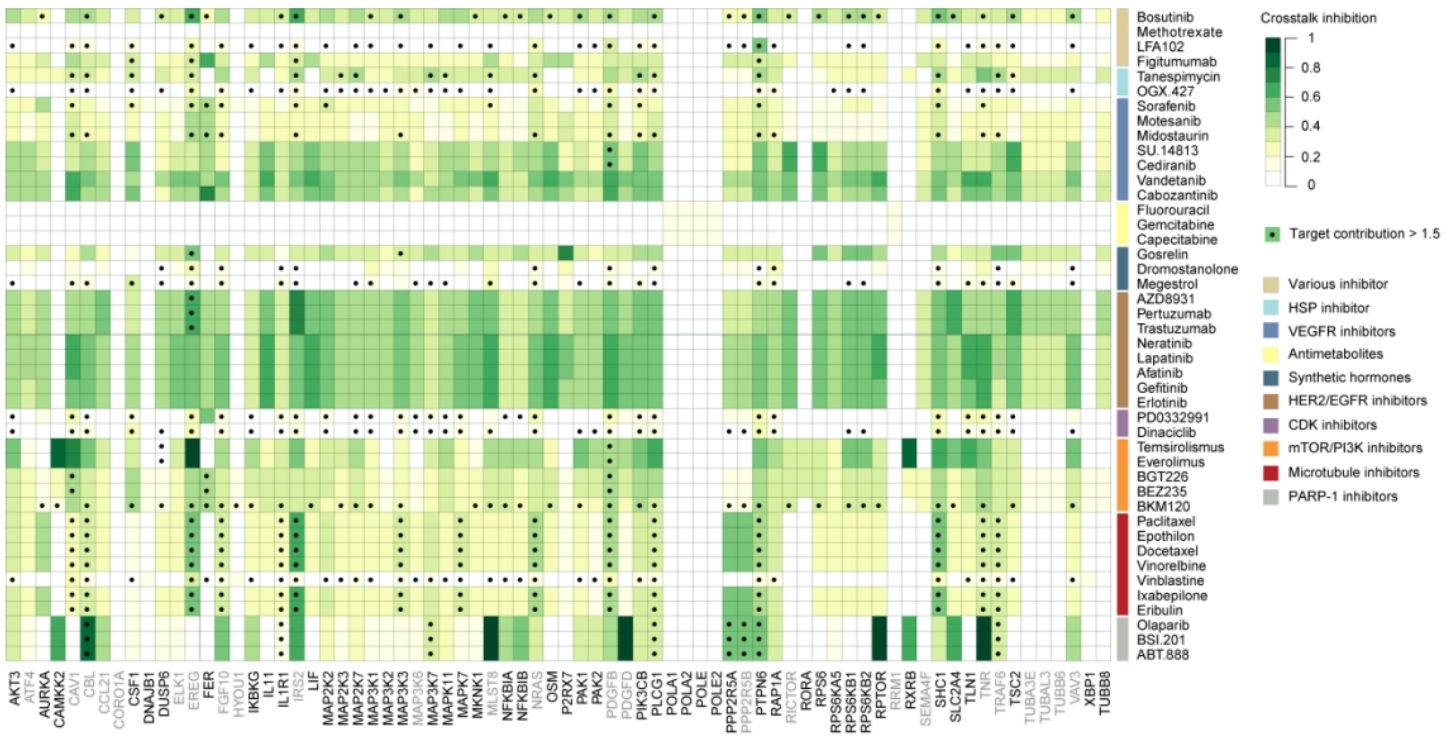


Figure 30. CI_C of target-drug combinations.

Crosstalk inhibition determined between novel candidate drug targets and the breast cancer drugs, involving 74 candidates and 44 breast cancer drugs from 10 different drug classes. Dotted fields indicate target-drug combinations in which the candidate target increases the crosstalk inhibition by 50% or more (BCT ≥ 1.5). For candidates in grey no chemical modulator is known.

6.4 Validation of novel combinations

Starting from 169 drug-drug combinations having a CI_C higher than the average observed in random combinations (0.25), we filtered out the most promising ones taking into account the computed CI_C and the prediction or not of an additive/synergistic effect. We also intended to maximize de variability of drug classes involved in the combinations. On the other hand, target-drug combinations were filtered based on the CI_C and the benefit of the candidate target (BCT). Thus, 10 drug-drug and 13 target-drug combinations were selected for experimental validation (see Tables 8 and 9).

Id	Drug1 (class)	Drug2 (class)	CI_C	A/S
DC01	**Cabozantinib** (VEGFR inhibitor)	**Erlotinib** (EGFR inhibitor)	0.60	
DC02	**Cabozantinib** (VEGFR inhibitor)	**Raloxifene** (SERM)	0.88	
DC03	**Olaparib** (PARP1 inhibitor)	**Tanespimycin** (HSP90 inhibitor)	0.80	
DC04	**Olaparib** (PARP1 inhibitor)	**Dinaciclib** (CDK inhibitor)	0.50	
DC05	**Olaparib** (PARP1 inhibitor)	**Palbociclib** (CDK inhibitor)	0.34	✓
DC06	**Cabozantinib** (VEGFR inhibitor)	**Palbociclib** (CDK inhibitor)	0.57	
DC07	**Paclitaxel** (Microtubule modulator)	**Tanespimycin** (HSP90 inhibitor)	1.00	
DC08	**Paclitaxel** (Microtubule modulator)	**Midostaurin** (VEGFR inhibitor)	0.37	
DC09	**Cabozantinib** (VEGFR inhibitor)	**Trastuzumab** (HER2 inhibitor)	0.43	
DC10	**Figitumumab** (IGF1R inhibitor)	**Raloxifene** (SERM)	0.72	✓

Table 8. Drug-drug combinations selected for experimental validation.

For each combination, the drug classes are specified, as well as the crosstalk inhibition (CI_C) and a potential additive/synergistic effect (A/S).

Id	Target (Modulator)	Drug (class)	CI_C	B_{CT}
TDC01	**PTPN6** (Sodium stibogluconate)	**Olaparib** (PARP1 inhibitor)	0.52	1.9
TDC02	**PTPN6** (Sodium stibogluconate)	**Palbociclib** (CDK inhibitor)	0.30	4.7
TDC03	**MAP3K7** (5Z-7-Oxozeaenol)	**Olaparib** (PARP1 inhibitor)	0.39	1.8
TDC04	**MAP3K7** (5Z-7-Oxozeaenol)	**Tanespimycin** (HSP90 inhibitor)	0.59	1.5
TDC05	**MAP2K2** (SL 327)	**Motesanib** (VEGFR inhibitor)	0.57	1.2
TDC06	**PIK3CB** (TGX-221)	**Motesanib** (VEGFR inhibitor)	0.56	1.1
TDC07	**PIK3CB** (TGX-221)	**Palbociclib** (CDK inhibitor)	0.20	5.5
TDC08	**PIK3CB** (TGX-221)	**Tanespimycin** (HSP90 inhibitor)	0.53	1.6
TDC09	**MAPK7** (XMD8-92)	**Paclitaxel** (Microtubule modulator)	0.37	1.6
TDC10	**IL1R1** (Anakinra [Kineret®])	**Paclitaxel** (Microtubule modulator)	0.30	2.1
TDC11	**PPP2R5A** (Norcantharidin)	**Olaparib** (PARP1 inhibitor)	0.60	-
TDC12	**PTPN6** (Sodium stibogluconate)	**Cediranib** (VEGFR inhibitor)	0.60	1.2
TDC13	**RAP1A** (GGTI-298)	**Palbociclib** (CDK inhibitor)	0.17	3.4

Table 9. Target-drug combinations selected for experimental validation.

In addition to the candidate target modulator and the drug class, the crosstalk inhibition (CI_C) and benefit of the candidate (B_{CT}) are shown.

Similarly to previously described for single target validation (see Section 6.1), each one of the 23 combinations was tested in six different cell lines (4 cells lines representing main breast cancer subtypes, an osteosarcoma cell line and a non-tumorigenic cell line) which meant a total number of 138 experiments. Given the data for the 138 combinations we applied sigmoidal fitting to model the respective dose-response curves from which we determined the half-maximal inhibitory concentration (IC_{50}). We then measured the drug

combination index (DCI) (Chou and Talalay, 1984)that provides a quantitative measure of the extent of the drug interaction. The DCI score for a drug combination is calculated as follows:

$$DCI_x = \frac{C_{D1}}{IC_{x,D1}} + \frac{C_{D2}}{IC_{x,D2}}$$

C_{D1} and C_{D2} represent the concentration of drug one and drug two used in combination to achieve an effect X while $IC_{X,D1}$ and $IC_{X,D2}$ indicate the concentration of the single agents which are required to induce the same effect X (Figure 31). In general, DCI scores for:

- DCI < 1 indicates synergism
- DCI = 1 indicates additive effects
- DCI > 1 indicates antagonism

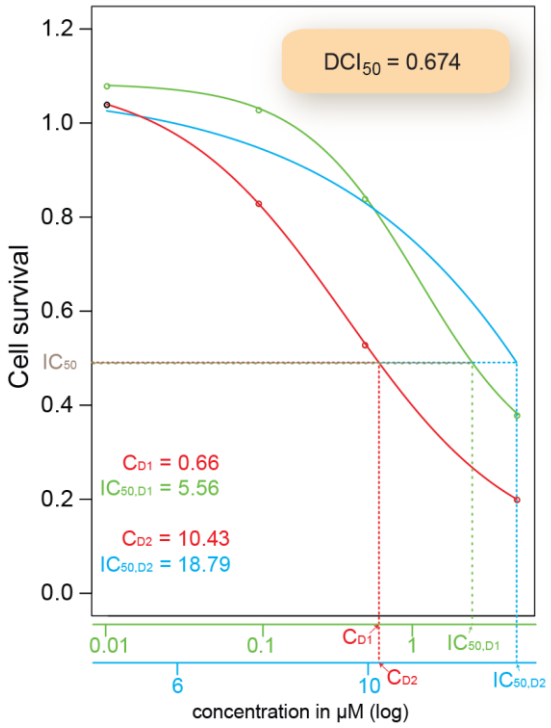


Figure 31. DCI calculation using Chou and Talalay method.

The DCI score measures the fractional shift between the combination doses ($C_{D1,X}$ and $C_{D2,X}$) and the single-agent inhibitory concentration ($IC_{X,1}$ and $IC_{X,2}$) for a given level of inhibition X.

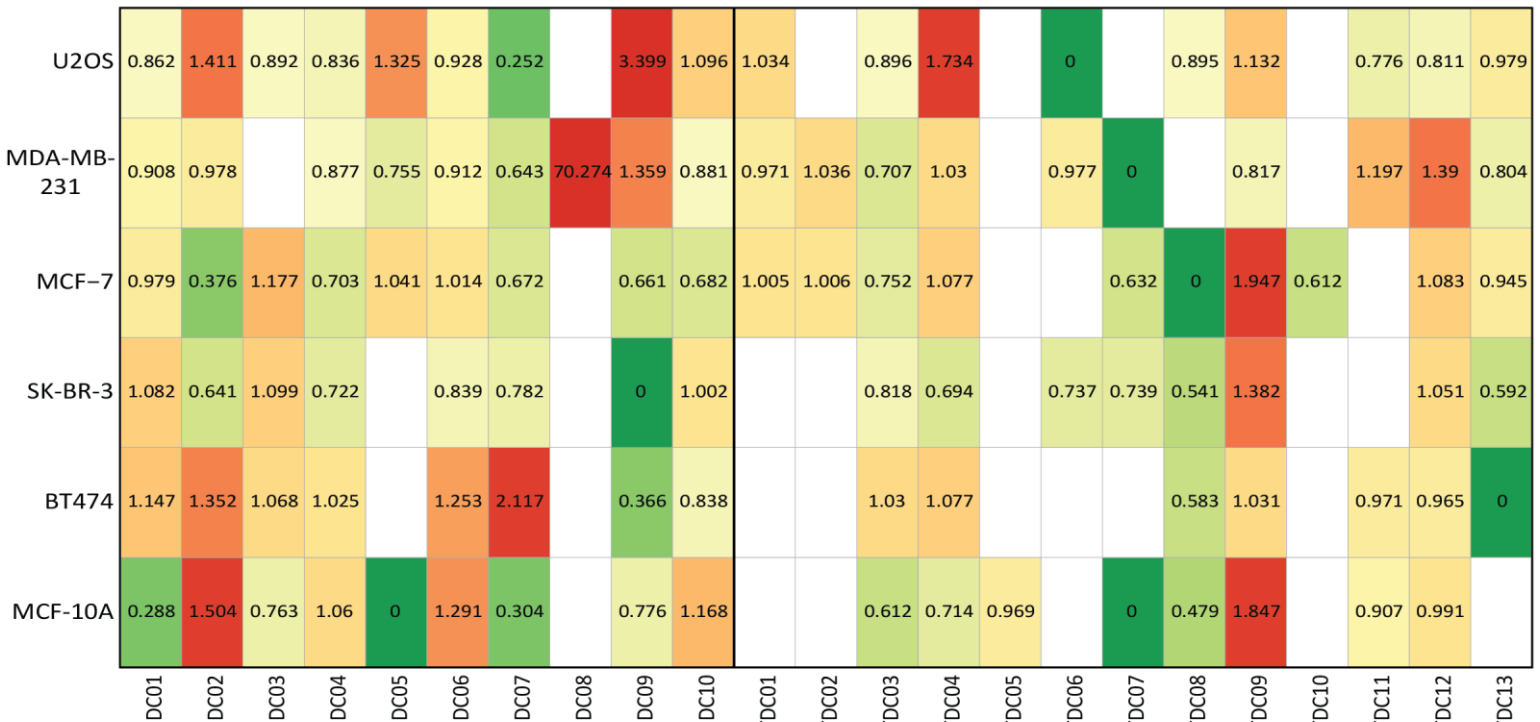


Figure 32. Drug-drug and target-drug combination results.

All DCI_{50} values computed from the experimental validation are shown. DCI_{50} of zero represents drug combinations in which the single agents do not reach the required IC_{50} while the combination does. Blank cells indicate combinations for which no DCI_{50} could be determined. Green represent synergism, yellow/orange cells represent an additive effect, while antagonistic combinations are red.

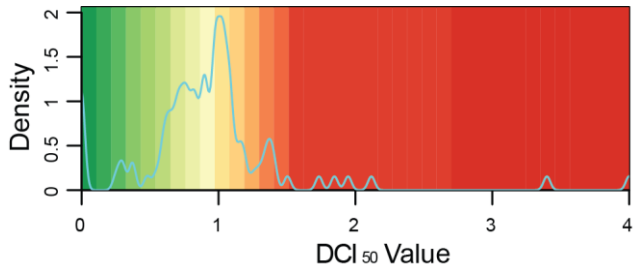


Figure 33. Color Key and Density Plot

Representation of the color scale based the DCI_{50} values on used in Figure 32. Density line shows a clear preponderance of synergistice/additive combinations.

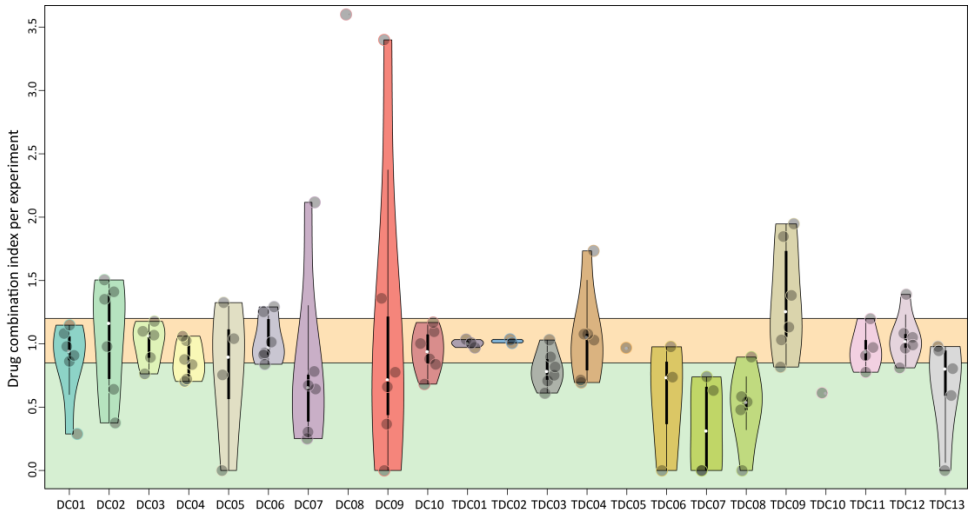


Figure 34. DCI_{50} indixes determined for the different drug combinations across the cell lines.

The green area indicates combinations that are considered to be strongly to slightly synergistic while combinations in the orange area are of additive nature.

In the Figures 32-35 we represent several different views of the results obtained from the 138 experiments carried out. Firstly, the DCI_{50} values computed in each cell line for each drug-drug and target-drug combinations are shown in Figure 32. In 33 combinations DCI_{50} was not determined, either because the IC_{50} could not be reached (20 combinations), or the induced response was far below the IC_{50} (one combination, DC08 in MCF-7) or no

dose-response curve could be modeled (12 combinations). We observed that a large number of the predicted combinations was actually moderately to highly synergistic in one or more cell lines. Considering DCI_{50} of the remaining individual experiments (one combination in one cell line) whose calculation was possible, the assays resulted in a higher porportion af synergistic results in comparisson to the antagonistic combinations. (Figures 32-33). The detailed percentages obtained are:

- 44 % of drug combinations are synergistic ($DCI_{50} \leq 0.85$)
- 39 % are of additive nature ($0.85 < DCI_{50} \leq 1.2$)
- 17 % are antagonistic ($DCI_{50} > 1.2$)

Apart from the general response results, the fact that we are testing combinations involving different drug classes in several BC subtypes allows multiple analyses. For example, we observed that, as expected, Raloxifene (an ER modulator) was more effective in MCF-7 (hormone receptor positive cell line) in both DC02 and DC10 combinations. Remarkably, the results obtained in DC02 highlighted VEGFR inhibitor Cabozantinib as a promising drug in combination with Raloxifene to treat ER+ breast cancer (DCI=0.376).

Another interesting example is DC09, where we combined:

① Trastuzumab: a monoclonal antibody that interferes with the HER2 receptor and it is used to treat HER2-positive breast cancer (Hudis, 2007).

② Cabozantinib: a small molecule inhibitor of VEGFR2 that was approved by the U.S. FDA in November 2012 for the treatment of medullary thyroid cancer and it is currently undergoing clinical trials for the treatment several cancers, including breast cancer (Smith et al., 2013).

The experiment performed indicated that this combination is highly synergistic in the HER2+ cell line (SK-BR-3, DCI=0) and in the triple positive cell line (BT474, DCI=0.366), and it is also synergistic in MCF-7 (DCI=0.661). Conversely, the combination of these two drugs showed an antagonistic effect in the triple negative BC cell line (MDA-MB-231, DCI=1.359) and in U2OS (DCI=3.399), as expected (Figure 32).

Figure 34 presents a combination specific overview, representing the distribution of DCI_{50} in the distinct cell lines in which we were able to determine it. We observed that some combinations, such as TDC09, were generally antagonistic, while others, such as TDC07 or TDC08, were predominantly additive or synergistic. In other cases as DC02 or DC09 there was a huge variability in the response of each cell line to this combinatory treatment, but a remarkable behavior was observed in specific BC cell lines.

Since the observed synergistic or antagonistic effects are likely to depend on the chosen dose or response level X, we determined the DCI for a number of IC values to be able to draw more meaningful conclusions from the DCI analysis. To this end, we computed the DCI score for response levels X = [20, 30, 40, 50, 60, 70, 80] which correspond to IC_{20} to IC_{80}. An example of the distribution of the DCI with respect to the different inhibition levels is shown in Figure 35 for combination DC04, and the full results of this analysis are shown in Appendix 10. In general, we observed that for some cell lines/combinations the DCI does vary largely depending on the inhibition level while for others there is less variance. For this reason, we also created a table to represent the average and median DCI values (including SD±) for each drug combination in each of the cell lines (Table 10). This table shows that, for

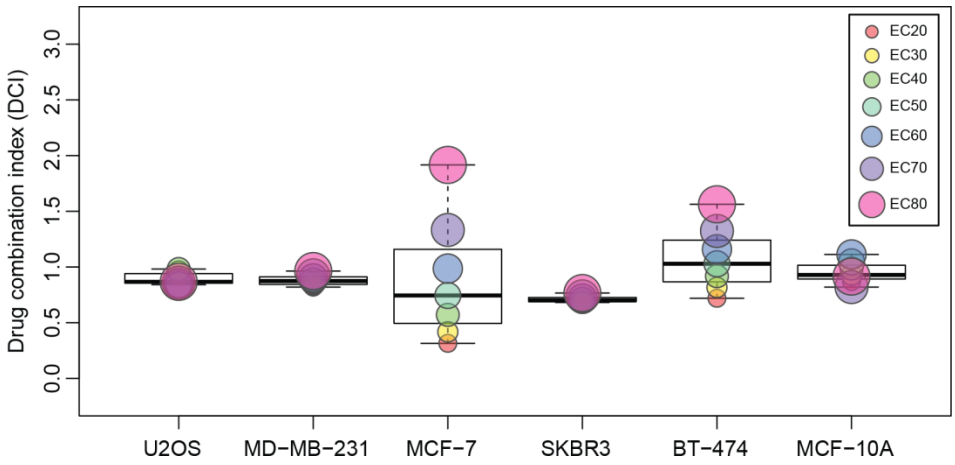


Figure 35. Distribution of DCI_{20}-DCI_{80} of drug-drug combination DC04.

The color and size of the nodes represents DCI with respect to different response level X = (20, 30, 40, 50, 60, 70, 80) determined for combination DC04.

DCI values	U2OS				MD-MB-231				MCF-7				SKBR3				BT-474				MCF-10A			
	mean	median	sd		mean	median	sd		mean	median	sd		mean	median	sd		mean	median	sd		mean	median	sd	
DC01	0,86	0,86	0,03	+	0,91	0,91	0,03	±	0,98	0,98	0,03	±	0,90	1,07	0,27	±	1,02	1,06	0,26	±	0,65	0,64	0,26	+++
DC02	1,26	1,41	0,25	--	0,98	0,98	0,11	±	0,46	0,38	0,30	+++	0,58	0,64	0,28	+++	1,52	1,35	0,53	--	1,49	1,50	0,06	---
DC03	0,95	0,90	0,51	+					1,04	1,05	0,10	±	1,11	1,10	0,01	±	0,84	0,98	0,42	±	0,55	0,70	0,46	+++
DC04	0,90	0,87	0,06	+	0,89	0,90	0,04	+	0,93	0,70	0,71	++	0,73	0,72	0,03	++	1,06	1,03	0,28	±	0,97	0,96	0,10	±
DC05	4,18E+14	1,33	1,02E+12	--	0,74	0,79	0,34	++	1,02	1,03	0,09	±	0,52	0,43	0,37	+++	0	0	0	+++++	0,15	0	0,26	+++++
DC06	0,98	0,93	0,12	±	0,92	0,91	0,03	±	1,01	1,01	0,01	±	0,85	0,84	0,14	++	1,26	1,25	0,15	--	1,08	1,29	0,50	--
DC07					4,62	2,45	6,00	---																
DC08					21,26	20,34	25,91	-----	0	0	0	+++++	1,50	1,50	2,12	---								
DC09	2,32	2,21	0,91	---	1,48	1,43	0,31	--	0,58	0,66	0,26	+++	107,66	0,97	214,04	±	0,59	0,37	0,60	+++	0,77	0,78	0,02	++
DC10	1,08	1,08	0,06	±	0,88	0,88	0,06	+	0,69	0,68	0,13	++	1,00	1,00	0,04	±	0,84	0,84	0,03	++	1,22	1,17	0,37	+
TDC01	1,42E+04	1,03	3,48E+04	±	0,84	0,97	0,37	±	1,01	1,02	0,04	±	1,26	1,26	0,12	--					1,18	1,18	0,06	-
TDC02					1,04	1,04	0,06	±	1,01	1,01	0,02	±	3,00	3,00	0,00	---	0,50	0,50	0,70	+++	0	0	0	+++++
TDC03	0,95	0,90	0,25	+	0,76	0,74	0,05	++	0,81	0,79	0,21	++	0,88	0,91	0,17	±	1,03	1,03	0,02	±	0,64	0,64	0,07	+++
TDC04	1,77	1,73	0,13	---	1,04	1,03	0,08	±	1,26	1,08	0,38	±	0,82	0,79	0,11	++	1,28	1,08	0,27	±	0,77	0,71	0,10	++
TDC05					0,73	0,72	0,12	++	1,53	1,53	1,75	---					0,80	0,80		++	1,19	0,99	0,56	±
TDC06	0,38	0,54	0,36	+++	0,97	0,98	0,00	±	121,18	1,29	208,00	--	0,80	0,79	0,09	++	0,94	0,94	0,02	±	0,81	0,81	0,09	++
TDC07	0,00	0,00	0,00	+++++	0,22	0,00	0,37	+++++	0,78	0,79	0,51	++	0,60	0,72	0,30	++	0,85	0,84	0,12	++	0,32	0	0,50	+++++
TDC08	0,97	0,90	0,20	++	0,98	0,98	0,01	±	0,92	0,46	1,29	+++	0,39	0,53	0,39	+++	0,62	0,58	0,53	+++	0,51	0,50	0,46	+++
TDC09					8,72	2,23	17,79	---	40,74	3,57	82,41	----												
TDC10									0,18	0,06	0,27	+++++												
TDC11	0,85	0,84	0,09	++	1,02	1,08	0,51	±									1,02	1,01	0,08	±	1,01	1,02	0,10	±
TDC12	0,81	0,81	0,02	++	1,40	1,39	0,20	--	1,08	1,08	0,02	±	1,05	1,05	0,01	±	0,97	0,97	0,02	±	0,99	0,99	0,02	±
TDC13	1,03	0,98	0,17	±	0,86	0,80	0,29	++	1,04	0,94	0,22	+	0,76	0,59	0,67	+++	0,60	0,82	0,58	++	0,58	0,85	0,51	++

Table 10. Overview of DCI values.

Average and median DCI values (including SD±) for each drug combination in each of the cell lines. Synergism (DCI ≤ 0.85), additive effect ($0.85 < DCI_{50} \leq 1.2$) or antagonism ($DCI_{50} > 1.2$) are indicated.

instance, the average and median DCI for DC01 in U2OS is 0.87 and 0.86, respectively, with a standard deviation of 0.02 indicating slight synergism between the two drugs independently from the inhibition level X. In turn, we detected a high standard deviation in the DCI average of DC04 in MCF-7 (0.93±0.71). In that case we can observe a correlation between the DCI and the inhibition level X (Figure 35): for smaller inhibition levels, such as IC_{20} - IC_{60}, the combination exhibits a synergistic or additive effect while for IC_{70} - IC_{80} results in very strong antagonism.

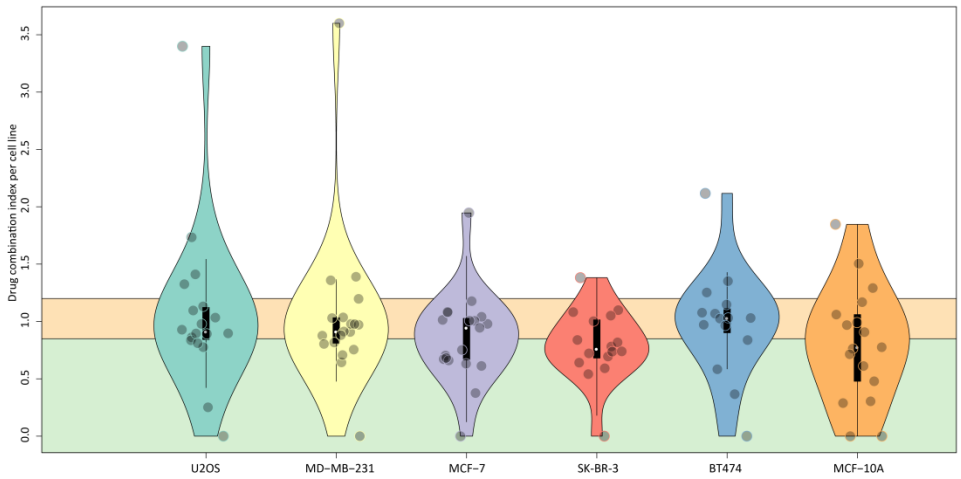


Figure 36. DCI distribution in each cell line.

The green area indicates combinations that are considered to be strongly to slightly synergistic while combinations in the orange area are of additive nature.

In Figure 36 we represent the distribution of DCIs per cell line. Although there are no big differences in the response among the cell lines, we can obsevre that MCF-7 and SK-BR-3 are slighty more sensitive to the drug combinations tested while BT474 is more resitant to the compounds used in combination. When analyzing the results it is important to discuss the behavior of the MCF-10A cell line. Our initial objective was to use this cell line as a control to measure the toxicity of the drugs in non-tumorigenic cells. However, when considering the outcomes of combinations and single candidate targets we cannot observe a difference between cell growth inhibition within cancer cells and MCF-10A. Although in many combinations we see the expected non-response of MCF-10A to drugs, there are a number of cases in which the

MCF-10A cells are actually more sensitive to a drug or a drug combination with respect one or more breast cancer cell lines (for example in DC05, TDC02, TDC07, …). For instance, Erlotinib and Cabozantinib, one being approved and the other one an experimental drug for breast cancer, inhibit cell growth in MCF-10A to a larger extent than actually in the five cancer cell lines. Similar observations have been made in other experimental studies where proliferation of MCF-10A cells was significantly suppressed by an anticancer drug (Hsieh et al., 2005). On the other hand, many other studies, such as the NCI60, which perform drug sensitivity assays in panel of cell lines don't use any control cell lines. Thus, we should be cautious when analyzing results from this cell line.

Effect of novel targets in combination

Irrespective of the inhibitory effects of single targets (see Section 6.2), we also assessed whether the candidate targets are involved in synergistic/antagonistic drug combinations. Table 11 shows that seven out of eight targets tested (PTPN6, MAP3K7, PIK3CB, MAPK7, IL1R1, PPP2R5A and RAP1A) showed synergism in at least one combination in any cancer cell line, while combinations involving MAP2K2 candidate targets where antagonistic. Interestingly, for some of the candidates without inhibitory effect observed that in combination they exhibit additive effects. For instance, a combination involving PTPN6 (TDC02 in U2OS) exhibited a synergistic behavior. Furthermore, we obtained data from only one combination involving IL1R1 (which did not lead to cell growth inhibition in none of the cell lines) we had only data from one combination (TDC10) in one cell line MCF-7). However, the combination showed a strong and consistent synergistic effect in this cell line.

Regarding to the contribution of MAP3K7 in combination with Olaparib (TDC03), we observed consistent synergistic or nearly additive effects across all the cell lines. Hence, MAP3K7 is a very promising target. On the one hand, it is itself an effective target, since IC50 can be reached in all the cell lines as well. On the other hand we observed that it is also a beneficial combinatorial target,

Target	Combination	Drug	U2OS DCI50		MD-MB-231 DCI50		MCF-7 DCI50		SKBR3 DCI50		BT-474 DCI50		MCF-10A DCI50	
PTPN6	TDC01	Olaparib	1,034	±	0,971	±	1,005	±						
PTPN6	TDC02	PD-0332991			1,036	±	1,006	±						
PTPN6	TDC12	Cediranib	0,797	++	1,34	--	1,083	±	1,005	±	0,953	±	0,99	±
MAP3K7	TDC03	Olaparib	0,896	±	0,707	++	0,752	++	0,818	++	1,030	±	0,612	+++
MAP3K7	TDC04	Tanespimycin	1,734	---	1,030	±	1,077	±	0,694	+++	1,077	±	0,714	++
MAP2K2	TDC05	Motesanib											0,969	±
PIK3CB	TDC06	Motesanib	0	+++++					0,737	++				
PIK3CB	TDC07	PD-0332991			0	+++++	0,632	+++	0,739	++			0	+++++
PIK3CB	TDC08	Tanespimycin	0,895	+			0,000	+++++	0,541	+++	0,583	+++	0,479	+++
MAPK7	TDC09	Paclitaxel			1,48	---	3,567	----						
IL1R1	TDC10	Paclitaxel					0,085	+++++						
PPP2R5A	TDC11	Olaparib	0,776	++	1,197	-					0,971	±	0,907	±
RAP1A	TDC13	PD-0332991	0,979	±	0,804	++	0,945	±	0,592	+++	0	+++++		

Table 11. Effects of novel BC targets in combination with BC drugs.

DCI_{50} scores for target-drug interaction only considering a cell growth inhibition of 50%. Blank cells indicate combinations where no IC_{50} could have been determined.

i.e., its combination with another breast cancer drugs, such as Olaparib, is synergistic (inhibition level of 50%).

Therefore, as we discuss later on, we can conclude that the proposed network biology approach is a robust strategy to identify novel breast cancer drug targets which are effective either alone or in combination with breast cancer drugs, as well as to identify novel promising drug combinations.

CHAPTER

III

GENERAL DISCUSSION

General Discussion

Proteins are essential in living organisms, as they participate in virtually every process within cells. Besides, the function of a protein is essentially performed due to its interaction with other proteins, indicating that protein-protein interactions (PPIs) play a key role in many cellular processes. Thus, the study of PPIs is becoming increasingly important in our effort to understand human diseases, such as cancer. Furthermore, if we are able to take a higher-level perspective and investigate the global network of PPIs we will improve the quality and the coverage of the research. Therefore, the identification and characterization of PPIs and their networks are essential to understand the mechanisms of biological processes.

The work presented in this thesis uses promising network-level applications such as identifying new disease genes; identifying disease-related sub-networks; and identifying new genes related to a specific cellular process (i.e. DNA repair). Moreover, we apply a network-based analysis in a more focused pharmacological area, i.e., target drug discovery and drug combinations prediction.

Enlargement of BC and CRC interactomes

Cancer is an extremely complex disease and innumerable proteins are involved in its development. Large genomic, proteomic and transcriptomic assays that have been and are being performed have contributed to the identification of many of these players. In most types of cancer, such as BC or CRC, the main pathogenic pathways that are related to their progression are being unveiled and studied. Nonetheless, there is still a long way to go in cancer research: on the one hand there are many unknown BC and CRC causing genes that remain to be discovered; on the other hand, we are unaware of the relevance, the role and the mechanisms associated to the carcinogenesis effect in most BC and

CRC associated genes. We hypothesize that an appropriate approach to solve and fill these gaps is to study protein-protein interactions, particularly using a network-based approach.

First of all we performed systematic matrix-based Y2H screens to identify novel interactions between genes that cause breast or colorectal cancers and genes associated to these cancers but whose role in the disease is unknown. The selection of drivers or causing cancer genes was similar in both studies: we chose well-established and characterized genes in BC and CRC development. However, some differences among the BC and CRC approaches need to be discussed. We followed two different strategies to select the associated genes, even though the objective in both studies was to place the candidate genes in the interactome. In the first approach (BC), the prioritization of the genes was carried out based on the network level. Our goal here was to opt for those genes whose network distance to the causative genes was higher. We discarded genes that were either interacting or that had one intermediate interactor to a causative gene. This way, we maximized to 100% the chance of detecting novel interactions and excluded the possibility to detect already reported interactions. On the other hand, in the second Y2H matrix screening (CRC) we picked genes whose expression was coordinated with CRC drivers across a compendium of normal tissues and cell types. Furthermore, we included the genes located in the 9q22.32-31.1 region, which contains a susceptibility locus (CRCS9) involved in the development of known hereditary colorectal cancer syndromes. Our aim once more was to detect genes in this region that interact with CRC drivers. These differences in final candidate genes selections do not affect the biological relevance of the new interaction data reported, as discussed later on. However, they are reflected in the percentage of interactions detected: although we were testing a higher number of interactions in the BC study (2,376 BC vs 1890 CRC), the approach used for candidate selection lead to less positive interactions (728 vs 1029) (see Table 12).

Secondly, we performed library-based Y2H screens to identify novel proteins that may relate to the development of BC and CRC. Again, we followed two different strategies to select the genes tested in the assay. In the BC library screening, we chose two of the three most relevant genes in BC therapeutic classification and prognosis: ER and ERBB2 (PGR was excluded because it

Approach	Interactions tested	Positive interactions	HC positive interactions
BC	2376	728 (31%)	491 (21%)
CRC	1890	1029 (54%)	595 (31%)

Table 12. Percentage of positive and high-confidence (HC) positive interactions in the two Y2H screenings performed.

showed self-activation activity in Y2H). We also picked nine genes that showed genetic predisposition to breast cancer. We came across 108 interactions (validated by pair-wise Y2H) among these 11 baits and 99 preys from the cDNA pool library. In using the Intogen database (Gundem et al., 2010), we verified that 76 of the 99 identified proteins had never been linked to breast cancer before. After analyzing these newly identified genes, we found that they are co-expressed in human tissues with their BC driver interaction partners, which validates the strategy followed. Furthermore, they are significantly overrepresented in tumor somatic mutations and, as the majority of the baits used for the library screen were genes whose mutations are associated to BC development, this feature confirms our hypothesis and strengthens their potential relevance in breast cancer.

So as to enlarge the CRC interactome, we selected six genes as baits based on gene expression profiles from normal mucosa, colorectal adenoma (benign tumor) and adenocarcinoma (Sabates-Bellver et al., 2007). In particular, we focused our study on the AXIN2, DLC1, PDGFRL, C9orf30, SFRP4 and SFRP2 genes, highly up-regulated in colorectal tumors compared to both normal mucosa and benign tumors. We detected and validated 27 novel interactions, involving the six baits tested and 20 novel proteins. Accordingly, these 20 proteins assemble a set of potential novel genes that could be relevant in CRC development, and should be taken into consideration in future colorectal cancer studies.

After merging the HC interactions from the Y2H matrix and library screens, we obtained two datasets of 599 and 622 novel interactions, which define our contribution to the enlargement of BC and CRC interactomes, respectively. Interestingly, two different analyses strengthen the quality of the data obtained.

First of all, we experimentally validated a random subset of HC interactions from the BC dataset using complementary techniques, and we obtained a 79% of validation via co-IP binding experiments in mammalian cells. These results strongly support the confidence of our Y2H interaction set. Secondly, we computationally evaluated whether the discovered interactions could really provide mechanistic details about the relationships between driver and candidate genes. We checked to see if each driver and its interacting candidates were generally involved in similar biological processes, in a procedure that we termed as 'functional coherence' of our HC interaction networks. The high coherence observed in the driver direct interactors from canonical pathways (similarity = 0.627, P-value < 0.0001), served as a positive control of the approach. Interestingly, both BC and CRC networks showed a high degree of biological process coherence (BC similarity = 0.347, CRC similarity = 0.321, P-value < 0.0001 in both networks). This high functional coherence motivates the future analyses using our novel Y2H interaction data. Furthermore, it indicates that our network-based approach is indeed a robust inference tool to gain insight into the underlying mechanisms of those proteins with previously unknown roles in BC, and it also offers a better understanding of the regulation between different proteins of complex biological systems.

BC interactome clustering

We next focused our attention on the data reported in the breast cancer interactome screening. In order to assess the relevance of these novel interactions in depth, it is necessary to study them in their biological context. Thus, we integrated them into the currently available human interaction data to build a comprehensive interactome associated with BC. We first merged our HC set of interactions with all the known interactors from the initial causing genes set, and we then extended this initial network to the next level, obtaining in this way a network of 11,226 interactions among 2,019 proteins, designated as BC-PIN.

It has been observed that proteins involved in the same cellular processes often interact with one another (von Mering et al., 2002), and several studies have shown that clustering an interactome is a powerful approach to identify functional modules (Wang et al., 2010). Therefore, we investigated the BC-PIN

structure to detect the presence of potential functional modules and we identified 178 modules in the BC-PIN, 146 of which showed a high degree of functional homogeneity, and 62 of them showed a significant enrichment of one or more GO biological process annotations. When we reanalyzed these functional modules excluding our 599 HC interactions, we observed a lower number of modules, as well as considerable less homogeneous and enriched groups. This result supports the idea that our study brings to light some unexplored areas of the BC-associated interactome, and significantly increases the understanding of known modules with more connections, which bolsters our results. In spite of only studying the modules of DNA repair to validate our data, we suggest that all the BC-PIN modularity data obtained (reported in Appendix 5) should give valuable information about the process in which the candidate genes are involved. Apart from the mentioned DNA repair function, we are also able to associate some of our candidate genes to a variety of processes such as apoptosis, signal transduction, transcription regulation, mRNA processing, protein folding, or ion transport, among others.

To sum up, the advantage of creating the BC-PIN network is that we are capable of interpreting and measuring the relevance of our novel interactions by taking into account their context and their weight in the global breast cancer interactome. Thus, we can conclude that based on the clustering analysis: (1) the contribution of our novel interaction data to BC interactome is remarkable, and (2) we are predicting specific cellular processes in which associated BC genes are involved.

Assessment of protein function predicted by interactome clustering

We merged all the modules that were homogeneous for DNA repair function to build a resulting DNA repair sub-network. To further investigate the relationships of genes present in our sub-network to DNA damage repair, we focused on 15 candidate genes that were either present in DNA repair modules or linking two or more of them. Interestingly, only one (RNF20) of these genes has been already described as a modulator of the DNA damage response, but a potential relationship with this process has never been established for the 14 remaining genes. Our main goal in this part of the project was to test if these

genes are indeed associated to the response to DNA damage. For this reason, we overexpressed and/or inhibited our genes of interest and we performed clonogenic survival and foci formation assays after cell irradiation. Overall, we detected a possible association of six genes (RNF20, SERPINB5, SNAIL1, FAM84B, MTA3 and IL24) to DNA damage response.

It is far from obvious that we can't expect all the predictions to be correct as we assigned the function(s) enriched in the cluster to all proteins with unknown

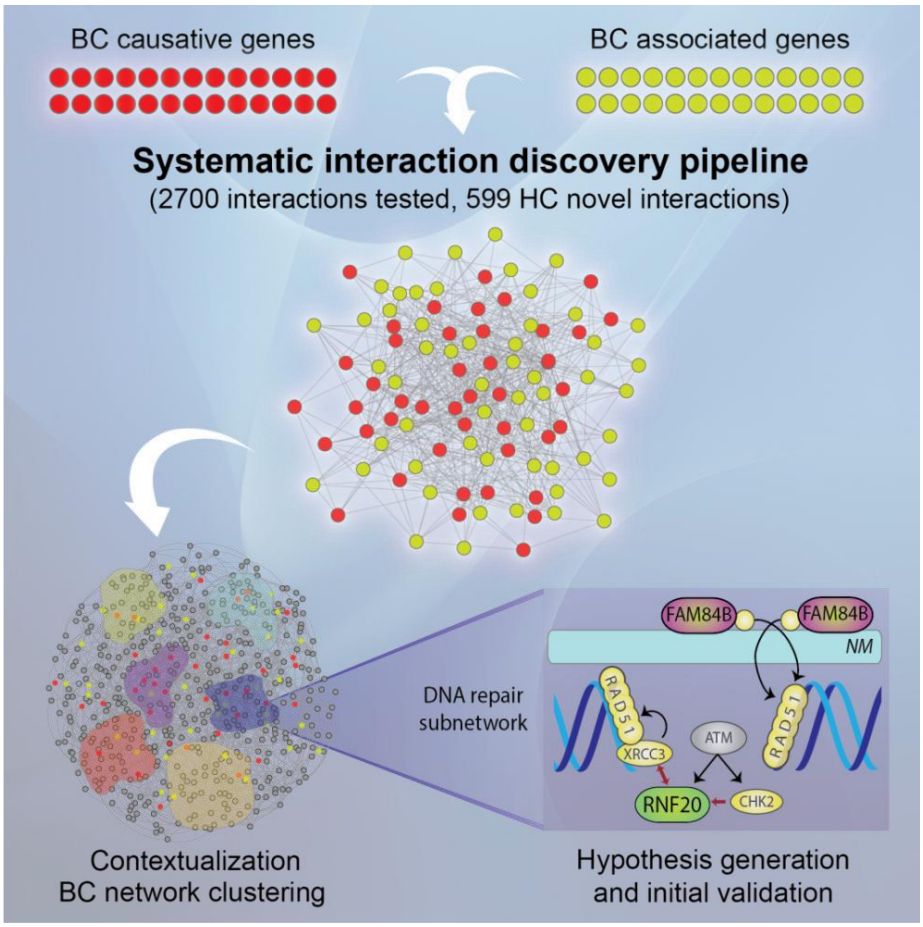


Figure 37. From interactome characterization to protein function prediction.

Starting from a selection of causative and associated BC genes, we followed three steps: (1) Systematic interaction discovery (2) Contextualization in the BC interactome, and (3) hypothesis generation and initial validation.

functions. These results indicate that, although more experiments need to be

performed to prove that those 15 genes play a pivotal role in DNA repair or in cell response to DNA damage, our module-assisted method is a reliable approach for protein function prediction.

We studied more deeply two genes that showed consistent positive results in the experiments performed: RNF20 and FAM84B. Regarding RNF20, recent studies have already demonstrated a role in recruitment of repair proteins upon ATM phosphorylation. Yet, we propose an additional role based on two novel interactions with XRCC3 and CHK2 reported in our Y2H screenings. We hypothesize that RNF20 may foster DNA repair mechanisms by favoring the XRCC3 binding to RAD51 to subsequently facilitate the formation of RAD51 nucleoprotein filaments that mediate HR repair. Furthermore, the association of RNF20 with CHK2 could also act as a linking mechanism for the ATM-CHK2 signaling, ensuring the DNA damage repair. For FAM84B, a protein whose function is unknown, our findings suggested that up-regulation of FAM84B in breast cells could promote early tumorigenesis by altering DNA repair mechanisms via RAD51 stabilization (Figure 37). Notwithstanding, considerable work needs to be achieved to fully elucidate the relevance of FAM84B in DNA damage repair.

In conclusion, the prediction of protein function via interactome clustering cannot be a substitute of a lab experiment, yet it provides a reliable reference of the cell function for these candidate proteins with unknown function.

A network biology approach to predict novel BC drug targets and drug combinations

In the second study we took into account functional redundancy and pathway crosstalk to predict novel drug targets and drug combinations for breast cancer.

Starting from list of 59 breast cancer driver genes compiled in the first study, we generated a breast cancer specific interaction network and then we identified proteins of the network that are targeted by current breast cancer therapeutics. The next step was to select the candidate proteins which are involved in similar biological processes as these current BC drug targets (210 candidate targets), and we finally ranked the identified candidate targets

according to their topological network similarity to BC targets. Our goal was to distinguish candidates resembling not only the functional but also the topological characteristics of breast cancer targets. After considering also literature evidence and the availability of chemical modulators we generated a final list of 54 putative promising novel BC targets.

Before performing traditional experimental validation, we analyzed the activity profiles of those chemical modulators associated with our 54 candidates that have been tested the NCI-60 cell line panel. Strikingly, the inhibition of putative novel drug candidate targets showed an increased activity against breast cancer cells lines, in comparison either to a random set of NCI compounds and also to a random set of proteins from the BC network. Furthermore, the activity data of the predicted targets was similar to the activity of the approved and experimental breast cancer drugs. Thus, the evaluation of the NCI60 activity profiles emphasizes the potential functional relevance of the candidate drug targets with respect to inhibit cell growth and to kill breast cancer cells.

We experimentally tested by MTT assays a selected set of eight candidate drug targets. Note that we carried out the assay in four different BC cell lines that were representing the main BC subtypes, in an osteosarcoma cell line and in a non-tumorigenic cell line. Two predicted targets did not show any effect (IL1R1 and PTPN6), and MAP2K2 inhibition affected mainly the non-tumorigenic cell line. Concerning PPP2R5A target, although we reached almost IC_{80} in four cell lines, the doses required were extremely high, and the most relevant effect was observed in the osteosarcoma cell line. Interestingly, the inhibition of four targets led to a relevant inhibition of several BC cell lines. While PIK3CB inhibition was able to affect 50-60% cell survival at most, the effect of the drugs targeting MAP3K7, MAPK7 and RAP1A was much higher (80%). MAP3K7 modulation, for example, consistently reduced cell growth by 80% in all the cell lines (including the normal breast cell line, MCF-10A). RAP1A, on the contrary was able to highly affect all cancer cells but it had no effect in MCF-10A. In conclusion, the inhibition of four out of the eight targets tested yielded a clear and consistent effect in the survival of breast

cancer cells, while other two were slightly inhibiting the grow of some subtypes of BC cells.

Next, we wanted to evaluate the applicability of the crosstalk inhibition metric for inferring novel breast cancer drug combinations. For this reason, we developed a computational approach for inferring pairs of current BC drugs combinations regarding drug resistance by taking functional redundancy and pathway crosstalk into account. Overall, we predicted 413 novel combinations that either exceeded the crosstalk inhibition threshold or showed a synergistic/additive behavior.

The next step of our strategy was to combine the two previous approaches: novel target identification and novel drug combination prediction. Thus, we exploited the potential pathway crosstalk inhibition (in a similar manner as described for drug-drug combinations) to simulate the impact of inhibiting novel candidate targets previously predicted in combination with breast cancer drugs. Overall, 1,108 combinations exceed the CI_C threshold of drug combinations in use (0.34) and 74 candidates were potentially beneficial for diminishing pathway crosstalk with at least one of the breast cancer drugs.

We carried out MTT assays to test a selected subset of 10 drug-drug combinations (DC01-DC10) and 13 novel target-drug combinations (TDC01-TDC13) predicted as potentially synergistic and/or with a significant crosstalk inhibition. A complete and systematic analysis of results from the tested combinations was performed. We assessed the DCI_{50} of each combination, and we further studied their behavior across the six cell lines used. The experiments indicate that considering DCI_{50} 44% of drug combinations interact in a synergistic manner with each other, 39% are additive and 17% are antagonistic. Apart from this global analysis, several examples previously discussed (see Section 6.4) indicate that our MTT experiments are providing relevant information about promising drug combinations in specific BC subtypes. We also computed DCI values for response levels from 20% to 80%, and all this relevant (which is available in Appendix 10) represents a complete view of the interaction behavior of the compounds tested in combination.

The last analysis of the MTT results was to assess the additive/synergistic effect of the novel targets in combination, irrespectively of the inhibitory effects of these single targets. Note that seven out of eight targets showed synergism in at least one combination. These results strongly validate the strategy followed to predict putative BC targets as only MAP2K2 target was not showing a significant effect neither alone nor in combination.

To sum up, the contribution of the second chapter of this thesis is: (1) the identification of novel breast cancer drug targets from the interactome providing promising individual or combinatorial inhibition with breast cancer agents. (2) The development of a method to quantify crosstalk between pathways and determining potential inhibition of this crosstalk by combined inhibition. (3) The experimental validation of a selected set of combinations predicted by this method demonstrating, on the one hand, the effect of combined inhibition on tumorigenic processes and on the other hand, the applicability of our network biology approach for cancer treatment. In conclusion, this network biology approach is a robust strategy that can be applied to other cancer types and diseases.

CHAPTER

IV

CONCLUSIONS

In the first part of this work, we show how a combination of interaction discovery experiments and the computational analyses of diverse biological data can provide further evidence for potential causative/susceptibility genes related to breast and colorectal cancers, suggesting novel hypotheses as to their molecular functions. We provide insights into the disruption of DNA damage response and repair mechanisms as a result of BC-related interaction re-wiring, albeit further studies have to be carried out to obtain a deeper mechanistic understanding.

Next, we proposed and experimentally validated a network biology approach to identify novel breast cancer drug targets which are effective either alone or in combination with breast cancer drugs. Another computational approach for inferring drug combinations regarding drug resistance by taking functional redundancy and pathway crosstalk into account was also developed and validated.

In conclusion, network-based strategies offer a global perspective to explore the molecular mechanisms underlying complex disease beyond individual genes and proteins. Therefore, the strategies presented in this thesis can be easily applied to other cancer types and diseases as well as for studying more specific biological questions related to breast cancer.

CHAPTER

V

MATERIALS AND METHODS

I conducted all the experimental methods described in this chapter. The computational methods, on the contrary, were performed by Dr Andreas Zanzoni, Dr Samira Jaeger and Miquel Duran, from my lab.

7. Experimental Methods

7.1 Subcloning of human cDNAs into Y2H plasmids (BC study)

61 clones from the human ORFeome v1.1 (Lamesch et al., 2007), 25 from Life Technologies Ultimate™ ORF Clones (Liang et al., 2004) and six from the Human ORFeome Collaboration Clone were obtained as a Gateway® cloning adapted plasmids and sequence verified. The 11 cDNA clones from human verified full length cDNA repository (ImaGenes) were amplified by PCR with primers containing additional nucleotides (CACC). These cDNAs were cloned into pENTR™/D-TOPO® vector (pENTR Directional TOPO cloning kit; Life Technologies) and sequenced.

All 103 ORFs were individually transferred into Y2H destination vectors by Gateway© recombinational cloning (ProQuest System, Life Technologies). Driver genes were cloned into pDEST32 to generate bait plasmids and associated genes were cloned into pDEST22 to obtain prey plasmids.

7.2 Subcloning of human cDNAs into Y2H plasmids (CRC study)

55 clones from the human ORFeome v1.1 (Lamesch et al., 2007), 27 from Life Technologies Ultimate™ ORF Clones (Liang et al., 2004) and 5 from the Human ORFeome Collaboration Clone were obtained as a Gateway® cloning adapted plasmids and sequence verified. Three cDNA clones from human verified full length cDNA repository (ImaGenes) were amplified by PCR with primers containing additional nucleotides (CACC). These cDNAs were cloned

into a pENTR™/D-TOPO® vector (pENTR Directional TOPO cloning kit; Life Technologies) and sequenced.

All 90 ORFs were individually transferred into Y2H destination vectors by Gateway© recombinational cloning (ProQuest System, Life Technologies). Driver genes were cloned into pDEST32 to generate bait plasmids and associated genes were cloned into pDEST22 to obtain prey plasmids.

7.3 Polymerase Chain Reaction (PCR)

PCR amplification of cDNA clones from Imagenes was accomplished with primers specific for each gene containing additional nucleotides (CACC). Primers were designed with Clone Manager (Sci-Ed) and VectorNTI (Life Technologies) softwares A 50 μL reaction, containing 1x PCR Buffer, 0.3 mM of each dNTP , 1 mM MgSO4, 0.4 μM forward and reverse primers, 100 ng of template DNA, 1U of DNA polymerase, and autoclaved water to reach final volume, was prepared and amplified in an Mastercycler ep Gradient (Eppendorf) using the following cycling parameters for touchdown PCR: initial denaturation at 95°C for 2 min; 30 cycles of denaturation at 95°C for 30 seconds, annealing for 40 seconds and extension at 72°C (temperature of annealing and extension time were different in each reaction based on primers and genes amplified); final extension at 72°C for 7 min.

7.4 TOPO® Cloning

TOPO® TA Cloning (Life Technologies) was permormed to create Gateway® adapted Entry Clones. We set up the following reactions using the reagents in the order shown: 4 ul of fresh PCR product, 1 ul of salt solution and 1 ul of TOPO® Vector. The reaction was mixed gently, incubated for 5 minutes at room temperature and transformed in competent E. coli cells (see below).

7.5 Gateway® Cloning

This reaction was used to transfer our genes from a Gateway® entry clone to destination vector. The following components were added to a 1.5 ml tube at room temperature and mixed: 150 ng of entry clone (1-5ul), 150 ng destination vector, TE buffer, pH 8.0 to 8 μl. Then 2 μl of LR Clonase ™II enzyme mix were added to the reaction and mixed well by vortexing briefly twice. This

reaction was incubated at 25°C for 1 hour. We added 1 µl of Proteinase K, vortexed briefly, and we incubated it at 37°C for 10 minutes to terminate the reaction. The reaction was finally transformed in competent E. coli cells (see below).

7.6 Transformation in bacteria cells

DH5α competent cells were aliquoted in 25 µl of cell mix for each transformation into 1.5 ml microcentrifuge tubes on wet ice. 10-50 ng of DNA (or 5 ul from Gateway Cloning reaction) were then added to each reaction tube by mixing gently and incubated on ice for 20-30 minutes. After incubation, cells were heat-shocked for 45 seconds at 42ºC without shaking. Tubes were immediately transferred to ice for two more minutes and 250 µl of pre-warmed medium (LB or S.O.C Medium) were added to each transformation mix, and cells are then incubated shaking at 37ºC for one hour. Cells were then spinned and all the supernatant but 40 µl was removed. Pelleted cells were resuspended and plated on pre-warmed selective LB plates. We used the appropriate selection marker for the LB plates suited to each destination vector (typically 100 µg/ml ampicillin).

7.7 Y2H co-transformation screens

Bait and prey plasmids were pair-wise co-transformed into MaV203 yeast strain in 96-well arrays. (1) Preculture: we re-plated yeast strain from an YPD-agar plate onto a new fresh YPDagar plate and we incubated it for a few days O/N at 30ºC. We scraped some colonies from the fresh plated YPD-agar plate and inoculate them into 50ul YPD. We vortexed until it was well resupended and it was inoculated into 3ml YPD + glucose and incubated O/N at 30ºC shaking at 250rpm. (2) Culture: we inoculated some of the preculture in an Erlenmeyer flask with 28ml YPD + glucose, so that the O.D 600 wass ~ 0.1. We then incubated the culture at 30ºC for aprox. 4h, shaking at 250rpm, until the O.D600 was ~ 0.6. (3) Transformation: We put a 96w microtiter PCR plate on ice and 5ul (or the equivalent to 1ug) of plasmid DNA (either bait or prey) was addead into each well of the PCR plate. We then added 1.5 ul of carrier DNA into each well and sealed the plate with an adhesive film. We prepared MIX1 (1 mL 1M LiAcO, 0.5 mL TE 10X, 5 mL 2 M Sorbitol and sterile water up to 10

mL) and MIX2 (1.5 mL 1M LiAcO, 1.5 mL TE 10X, 10 mL PEG3350 60% and sterile water up to 15 mL) and kept them shaking until use. The yeast culture was centrifuged at 2000 rpm for 5', discarded the supernatant and we added 10mL of TE 1X and resuspended the pellet. We centrifuged it at 2000rpm for 5', discarded the supernatant, then we added 1100ul of MIX1 into the cell pellet and vortexed it very carefully. It was incubated at RT for 10' and we dispense 11ul of the yeast suspension into each well of the PCR plate. We added 58ul of MIX2 into each well, sealed and vortexed it and we incubated at 30ºC for 30'. We added 8ul of DMSO into each well, sealed it and we incubate the PCR plate at 42ºC for 7' using the PCR Cycler machine (heat shock step). Finally, with a 96 pin replicator, we took ~10ul of the cell prep from each well and stamped it onto a fresh SD2-agar OmniTray. The plate was incubated upside down at 30ºC for 3-4 days until colonies appeared.

Co-transformed plasmids were plated onto selective SD2 (lacking Leu and Trp amino acids) agar media and incubated for 48 hours at 30ºC to detect colony growth. Co-transformant arrays were then replica plated onto different selective media for interaction screening. To assay the activation of the HIS3 reporter gene, SD3 (lacking Leu,Trp,His) agar plates were supplemented with 12 to 100 mM of 3-aminotriazole (3AT, Sigma-Aldrich), being 50 mM 3AT the optimal concentration for positive HIS3 activation colonies. Similarly, we assayed the activation of the URA3 reporter gene by plating onto SD3 (lacking Leu,Trp,Uracil). Double reporter HIS3/URA3 activation was evaluated by SD4 (lacking Leu,Trp,His, Uracil) agar plates supplemented with 20 mM of 3AT and LacZ reporter gene was tested by the β-galactosidase assay on a nylon membrane placed onto a SD2 agar plate.

7.8 Y2H library screens

Y2H library screens were performed using an adult human brain cDNA prey library (ProQuest, Life Technologies). The overlapping transcriptome ratio between brain and breast is 0.93, according to the TissueInfo database (Skrabanek and Campagne, 2001). Yeast cells expressing individual baits were transformed with the cDNA library using the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl, 2007) and were screened onto selective agar media for reporter gene activation. We checked HIS3 and URA3 reporter gene

activation by 7 day incubation of transformed cells at 30ºC in selective agar media (HIS- and/or URA-). We picked up positive growing colonies and cultured them in prey selective liquid medium (lacking Trp). In each screen, approximately 5×105 auxotrophic transformants were tested on selective plates, obtaining 20-40 positive colonies in average. The prey plasmid DNA was extracted from the cultures and the bacterial transformation of each plasmid was carried out in order to enable DNA sequencing and subsequent gene identification by BLAST search. The preys identified by the library were tested with their respective baits for activation of reporter gene expression in co-transformation assays, in a similar procedure as explained above.

7.9 In vitro co-immunoprecipitation (co-IP) assays

For expression of FLAG and Myc-tagged fusions, cDNA fragments (identical to those in the Y2H assays) were subcloned into SF-TAP or pDEST-Myc vectors, respectively, using the Gateway® system (Life Technologies) and transfected into HEK293T cells. cDNA plasmids were transfected using Lipofectamine 2000® (Life Technologies) or polyethylenimine (PEI) (Polysciences). Two days after transfection, cells were harvested and lysed with lysis buffer [0.5% TX-100, 150mM NaCl, 1mM EDTA, 1mM MgCl2, 50mM Tris-HCl (pH 7.7) and complete protease inhibitor cocktail (Roche)].

Whole cell lysates were cleared by centrifugation for 20 minutes at 16,000 x g at 4ºC and then precleared by adding Dynabeads® Protein G (Life Technologies). After 30 minutes of rotation at 4ºC, we removed the beads by centrifugation at 16,000 x g at 4°C for 10 minutes and 1ug of the appropriate antibody was added to the lysate. After incubation for 1 hour at 4ºC on a rotating plate, 10 ul of Dynabeads® Protein G were added and incubated under rotation at 4ºC overnight. After washing three times with lysis buffer and using DynaMagTM® (Life Technologies), bound proteins were analysed by SDS-PAGE and immunoblotting with the corresponding antibody. The antibodies used were mouse anti-Myc monoclonal antibody (mAb) and rabbit anti-FLAG polyclonal antibody. For co-IP of endogenous proteins, the same protocol was followed using primary antibodies: rabbit anti-FAM84B polyclonal antibody and rabbit anti-RAD51 polyclonal antibody.

7.10 Double immunofluorescence and confocal microscopy

U2OS cells were harvested at 24 h post-transfection (with Myc or FLAG-tagged proteins), fixed with 4% paraformaldehyde in phosphate buffered saline (PBS) and permeabilized in 0.1% Triton X-100 in PBS. Cells were blocked with 1% BSA in PBS and reacted with the proper antibodies. The antibodies used were a rabbit anti-FLAG antibody (1:200 in PBS-BSA) and Alexa Fluor 488-labeled goat anti-rabbit IgG (Life Technologies), and a mouse anti-Myc monoclonal antibody (1:200 in PBS-BSA) and Alexa Fluor 568-labeled goat anti-mouse IgG antibody (Life Technologies). To stain the nuclei, Hoescht was added to the cells and incubated for another 5 min and they were subsequently mounted on slides using ProLong® Gold antifade reagent (Life Technologies, cat. P36930). Samples were observed using a Leica TCS SP2 confocal microscope.

7.11 RNA interference-mediated gene silencing

Cells were seeded and exposed to 100 nM of either specific or scrambled control small interfering (si)RNA, using Lipofectamine® RNAiMAX transfection reagent (Life Technologies) for 24-48 h. siRNA sequences were designed according to MitoCheck database (http://www.mitocheck.org/). Specific silencing was monitored with β-actin as a loading control probe using anti-β-actin (Abcam, cat. ab20272). siRNAs were purchased from Life Technologies

7.12 Clonogenic survival assays

Cells were plated in 60-mm Petri dish and transfected with Myc-tagged proteins or with siRNAs against the targets of interest (as described above). After 24-48h, cells were plated onto 60-mm dishes at different densities in duplicate, and were exposed to various IR doses at 1.62 Gy/min rate. After 14 days of incubation at 37ºC, the cultures were fixed and stained with Crystal Violet Solution (Sigma-Aldrich). Colony-forming efficiency was calculated as the average of triplicate experiments by counting the number of colonies containing more than 50 cells.

7.13 Foci formation assay

10^5 U2OS cells were seeded into 6-well plates containing a glass coverslip in each well and were transfected with Myc-tagged proteins. Cells were irradiated with 1 Gy X-rays. After treatment, cells were fixed in 4% PFA, washed with PBS, permeabilized in 0.1% Triton-X 100 in PBS and blocked with 1% BSA in PBS 30 min. Samples were then incubated with rabbit monoclonal anti-γH2AX (Millipore, cat. 05-636) and mouse anti-Myc monoclonal antibodies (Life Technologies, cat.13-2500) for 2 h. Subsequently, they were incubated for 1 h with Alexa Fluor 488-labeled goat anti-rabbit IgG (Life Technologies) and Alexa Fluor 568-labeled goat anti-mouse IgG antibodies (Life Technologies). Nuclei were stained with DAPI. Coverslips were then removed from the plate and mounted onto a glass slide using ProLong® Gold antifade reagent (Life Technologies, cat. P36930). Samples were observed with a Leica TCS SP2 confocal microscope.

7.14 Subcellular fractionation

After removing the medium, cells were rinsed with PBS and detached, pelleted by centrifugation in 15-ml tubes, transferred to 1.5-ml tubes, and washed with PBS. Cell pellets were suspended in buffer E [20 mM HEPES-KOH (pH 7.8), 5 mM potassium acetate, 0.3 mM MgCl2, 0.5 mM dithiothreitol and protease inhibitors]. After 10 min on ice, samples were homogenized using a douncer and centrifuged at 1500 x g for 5 min. Supernatants were transferred to fresh tubes and centrifuged at 16.000 x g for 45 min and the supernatant (cytosolic fraction) and pellet (organelles fraction) were stored on ice. Pellets from the first centrifugation were resuspended in buffer N [20 mM HEPES-KOH (pH 7.8), 5 mM potassium acetate, 0.3 mM MgCl2, 0.5 Mm dithiothreitol, 540 mM NaCl, 10% glycerol and protease inhibitors], after which nuclear and chromatin bound proteins were extracted by end-over-end rotation at 4ºC for 90 min. Samples were centrifuged for 15 min at 16,000 x g, and supernatants (nuclear fraction) were transferred to fresh tubes. Pellets were washed with buffer N and centrifuged briefly, and the supernatant was removed. Protein concentrations were determined, and samples were prepared for Western Blot

analysis. Mitochondrial cellular fraction was verified by anti-VDAC loading control probe, nuclear fraction by anti-DNMT1 and cytosolic fraction by anti-β-actin.

7.15 MTT assays

Prior to the experiments, the optimal cell density was assessed to ensure exponential growth of cells at days 3, 5 or 6 in 96-well microculture plates in conditions of low serum culture used for the experiments. To be more specific, the number of cells seeded per well were 2,500 MCF-7, 5,000 SK-BR-3, 5,000 BT474, 4,000 MDA-MB-231, 2,500 U2OS and 3,000 MCF-10A for 3-day experiments, and 1,000 MCF-7, 2,500 SK-BR-3, 2,500 BT474, 1,500 MDA-MB-231, 1,000 U2OS and 1,000 MCF-10A for 5 or 6-day experiments. Cells were seeded in 96-well plates at these densities in 100 μL per well and cultured overnight, after which the corresponding drug was added at 4 different doses and incubated at 37° C in a humidified atmosphere with 5% CO_2 for specified exposure intervals (3, 5 or 6 days) (Table 13). At that time, 20 μL of MTT (Sigma; 5 mg/mL of PBS) was added to each well, and the plates were incubated for 3 h at 37°C. Then, the supernatant was aspirated; 100 μL of DMSO was added into each well; and absorbance was measured at 570 nm.

In cases of combined treatment, both drugs were added simultaneously and cells were exposed continuously to the drugs. All experiments were performed in triplicates (single knon BC drugs) or quadruplicates (novel targets and novel combinations) and confirmed in at least three independent experiments, SDs were obtained. Results are expressed as percentage of cell survival, calculated as the absorbance of treated well/absorbance of untreated control × 100. Table 13 shows the details of each compound tested in the MTT assays (name, supplier and CAS number) as well as the experimental conditions used (solvent, stock solution prepared, the four different doses used for testing and the incubation time in which each drug was tested).

	Name	Supplier	CAS number	Solvent	Stock Solution (mM)	Doses Tested	Incubation time (days)
D01	**Cabozantinib (XL-184)**	Selleckchem	849217-68-1	DMSO	20	5 / 7,5 / 10 / 15 µM	3, 5
D02	**Dinaciclib (SCH727965)**	Selleckchem	779353-01-4	DMSO	1	0,5 / 5 / 10 / 50 nM	3
D03	**Erlotinib**	Selleckhem	183321-74-6	DMSO	5	0,5 / 1 / 7,5 / 15 µM	3
D04	**NVP-AEW541 (~Figitumumab)**	Selleckchem	475489-16-8	DMSO	20	5 / 7,5 / 10 / 15 µM	3
D05	**Trastuzumab (Herceptin)**	Farmacia	180288-69-1	water	20 mg/ml	5 / 10 / 20 / 40 µg/mL	3
D06	**Paclitaxel (Taxol)**	Selleckchem	33069-62-4	DMSO	20	0,1 / 1 / 7,5 / 15 µM	3
D07	**Midostaurin (PKC412)**	Tocris	120685-11-2	DMSO	10	0,1 / 0,5 / 1 /10 µM	3
D08	**Olaparib (AZD2281)**	Selleckchem	763113-22-0	DMSO	20	0,01 / 0,1 / 1 / 5 µM	3, 5, 6
D09	**PD-0332991 (Palbociclib)**	Selleckchem	571190-30-2	water	4	0,01 / 0,1 / 0,5 / 5 µM	3, 5, 6
D10	**Raloxifene**	Selleckchem	84449-90-1	DMSO	20	0,1 / 1 / 10 / 20 µM	3
D11	**Tanespimycin (17-AAG)**	Selleckchem	75747-14-7	DMSO	2	0,2 / 2 / 20 / 200 nM	3
D12	**Cediranib (AZD2171)**	Selleckchem	288383-20-0	DMSO	20	0,1 / 1 /10 / 30 µM	6

D13	**Motesanib (AMG-706)**	Selleckchem	453562-69-1	water	20	0,01 / 0,1 / 1 / 10 µM	3
D14	**Anakinra (Kineret)**	Sobi	-	water	9	0,1 / 1 / 5 / 10 µM	3
D15	**SL 327**	Selleckchem	305350-87-2	DMSO	10	0,1 / 1 / 10 / 20 µM	3
D16	**5Z-7-Oxozeaenol**	Tocris	66018-38-0	DMSO	10	0,01 / 0,1 / 1 / 10 µM	3
D17	**XMD8-92**	Tocris	1234480-50-2	DMSO	20	0,1 / 1 / 7,5 / 15 µM	3
D18	**TGX-221**	Selleckchem	663619-89-4	DMSO	20	0,1 / 1 / 5 / 20 µM	3
D19	**Norcantharidin**	Sigma	29745-04-8	DMSO	100	0,1 / 1 / 10 / 100 µM	3
D20	**Sodium stibogluconate**	Millipore	16037-91-5	water	2.2	1 / 10 / 50 / 100 µM	6
D21	**GGTI-298**	Sigma	1217457-86-7	DMSO	20	0,1 / 1 / 5 / 20 µM	6

Table 13. Experimental deConditions used to test drugs by MTT assay.

8. Computational Methods

8.1 Connectivity assessment

We built a human interactome fetching the most recent available data (June 2010) from the BioGRID, DIP, IntAct and MINT databases (Aranda et al., 2010; Breitkreutz et al., 2008; Ceol et al., 2010; Salwinski et al., 2004). We selected experimentally verified direct interactions and added those interactions described as binary according to the associated detection methods (Rual et al., 2005). We further extended the interactome including the HPRD dataset (Prasad et al., 2009) obtaining a human binary interactome consisting of 9709 interactions between 35707 proteins. We then evaluated the interconnectivity of breast cancer related genes in terms of average shortest path length. To assess the statistical significance of the connectivity measure, we defined a reference distribution of 10'000 instances of size equal to 59 (the number of breast cancer related genes) consisting of randomly picked proteins from the human binary interactome. We compared the breast cancer genes average shortest path length and the random set average shortest path length using the Mann-Whitney U-test.

8.2 Gene coexpression analysis

In order to assess the degree of coexpression between the novel interactions discovered after cDNA library screening, we downloaded human coexpression data from COXPRESdb v4 (Obayashi et al., 2013). Using full coexpression profiles of driver genes as a background, we performed a significance test analogous to gene set enrichment analysis (GSEA) (Subramanian et al., 2005),

where gene pairs were rank-ordered according to their coexpression correlation, and the overall high rank of pairs corresponding to our experimental interactions was tested. The R-Bioconductor phenoTest v.1.7 package was used for the computations.

8.3 Correlation in gene expression profiles

We used microarray data from (Su et al., 2004), a compendium of gene expression profiles from 73 normal tissue and cell types. As input, we used this compendium of gene expression profiles, a list of known disease-related genes (OMIM, COSMIC and literature based driver/causing genes) and a list of candidate disease genes. We then applied a mixture model in order to obtain correlation coefficients that were robust to the presence of noise and we fit the model using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The procedure computes gene expression correlation coefficient among disease genes and also between known and candidate disease genes. We defined that known disease gene and a candidate disease gene as co-expressed if their EM correlation coefficient was more than 2 and the probability of noise less than 2.

8.4 Functional coherence of the HC interaction network

Functional similarity between pairs of proteins was obtained using best-match average of G-SESAME semantic similarities among annotated GO biological processes (Du et al., 2009) [UniProt annotation file (April 30, 2013), including electronically inferred terms, GO ontology file (May 6, 2013), with 'cross-products' and 'has part' relationships removed]. For each driver and its corresponding interactors, we defined functional coherence as the average of pairwise functional similarities. The significance of this coherence was assessed after assigning for 10'000 times an equal number of randomly picked proteins from the human binary interactome. Analogously, we calculated a global functional coherence degree by simply averaging the functional similarities of

our experimental interactions. To supply compelling evidence that functional coherence around driver/causing genes can indeed infer mechanistic insights, we conducted the same analysis with driver direct interactors that are annotated in Reactome pathways (May 6, 2013; (Matthews et al., 2009).

8.5 Gene Ontology annotation

We used the human GO annotation extracted from the Entrez gene2go file (NCBI, June 2010) and assessed the statistical significance of GO term enrichment using the Fisher's exact test. We adjusted the P-values for multiple testing applying the Bonferroni correction.

8.6 Identification of functional modules within the BC-PIN

We applied the MCL algorithm (Enright et al., 2002) to identify cluster representing putative functional modules. Since the granularity of the clustering depends the inflation coefficient I, we ran MCL on the BC-PIN exploring a wide range of I (from 0.1 to 10.0 by steps of 0.1). We chose the value of I that maximized the number of functionally homogenous clusters, i.e. modules, containing at least 3 proteins. We evaluated the functional relatedness of modules in terms of GO homogeneity GH (Goh et al., 2007), defined as the maximum fraction of proteins in the same module that have the same GO terms from the biological process branch. For the GH computation, we required the 50% of the proteins to be present in the module to be annotated with at least one GO term. We then assessed the statistical significance of each homogeneous module comparing its GH to the mean GH of a reference distribution obtained by computing the GH for 10'000 random generated sets of the same size of the module. We picked proteins for the randomization from the human binary interactome.

8.7 Merging of functional modules based on their semantic similarity

We evaluated the relatedness between the GO annotations of functionally homogeneous modules, we computed the semantic similarity among all annotation pairs with the GoSemSim package in R (Yu et al., 2010) using the Wang similarity measure that proved to be more robust that other methods (Wang et al., 2007). In order to select the functions to be grouped together, we assessed the statistical significance of the similarity of each pair comparing its Wang measure to the one of a reference distribution of 10'000 random generated pairs of GO annotations. We kept those annotation pairs having a Bonferroni adjusted P-value < 0.05.

8.8 Assignment of functional domains to the interacting candidates

We defined two reference datasets, both downloaded/generated in December 2010: 1) DNA repair genes: 162 genes from the updated list of Wood et al 2005 (Wood et al., 2005) and 2) DNA damage response genes: 392 genes annotated with the Gene Ontology (Consortium, 2010a) annotation 'response to DNA damage stimulus' (GO:0006974) in UniprotKB (Consortium, 2010b).

Fifteen genes were selected for further investigation in the context of DNA repair/damage response. For each reference gene set, we fetched Pfam-A (Finn et al., 2010) and SMART (Letunic et al., 2009) domain assignment from the respective websites. We then checked if any of the candidate had at least one Pfam-A or SMART domain found in any gene of the two reference gene sets. For 7 genes, we found at least one domain present in at least one gene associate in DNA repair and in DNA damage response. The majority of the domains are involved in DNA-binding (RING, zf-C3HC4, ZnF_C2H2, PHD) and in mediating protein interactions (PWWP, WD40).

8.9 Statistical analysis

We initially looked for differences in clone counts at increasing irradiation doses for each gene compared to controls. For this purpose, we fitted a linear model to the logarithm of counts adjusted by date of experiment. To detect either increasing or decreasing trends across doses, we normalized counts first subtracting the initial number of clone counts vs. dose for each gene. Next, we subtracted the counts of the non-radiated experiments. A linear model was then fitted to the normalized log-counts in order to find statistically significant differences between control and gene counts across doses.

8.10 Drug target identification strategy

Our method for identifying novel drug targets is composed of three steps. First, we generated a breast cancer specific interaction network around genes involved in breast cancer-relevant processes. Starting from the gene products of the 59 driver/causing genes we grew a network by integrating all proteins that interact either directly or indirectly with any of the drivers in the human interaction network.

Using this protein network we then identified those proteins which occupy the same functional space, i.e., similar biological processes and pathways, as known therapeutic breast cancer targets. To this end, we computed the functional similarity between candidates and breast cancer drugs, by means of their targets, using their pathway and GO annotations. The similarity between two sets of pathways was determined by employing Jaccard similarity. Functional GO similarity between two proteins was calculated using a semantic similarity measure proposed by (Couto et al., 2007). Both pathway and GO similarity for biological process were then combined using the arithmetic mean. To identify proteins with significant functional similarity, we compared the determined similarities against a background distribution of pairwise similarities between breast cancer drugs and proteins from the breast cancer network. We considered all proteins with statistical significant functional similarity (P-value < 0.0015) as candidate drug targets.

In the last step, we prioritized candidate drug targets according to their topological similarity to approved breast cancer drug targets within the interaction network. For this purpose, we first examined different topological features that are thought to be important for successful drug targets (Zhang and Huan, 2010; Zhu et al., 2009), namely degree, cluster coefficient, betweenness centrality, 1N index and topological network similarity (Erten et al., 2011). In addition, we also considered a protein's relationships to breast cancer drivers and approved drug targets with respect to average and shortest distance as well as its participation in important cancer driver pathways. To analyze which features are essential for distinguishing drug targets from non drug targets, we compared these ten features among five different protein groups, namely (1) approved and (2) experimental breast cancer drug targets, (3) other (non breast cancer) drug targets, (4) driver proteins and (5) remaining proteins. Based on this analysis we excluded cluster coefficient and topological network similarity as no significant differences could be detected for these features. We determined the remaining eight features for each approved and candidate drug target. Next, we ranked each candidate according to its distance to the median of a feature determined for approved targets. This resulted in eight rankings which then were fused to yield an overall rank for each candidate target. For combining the individual rankings we tested three strategies, (i) average and (ii) median rank across all ranking as well as(iii) rank order statistics (Aerts et al., 2006), which have been evaluated in a cross-validation setting for approved, experimental, non-breast cancer drug targets and randomly picked proteins. For each of those proteins, we determined topological properties and distance to the approved targets to assess its rank among all proteins of the network. Finally, we chose the average ranking strategy to rank the candidate drug targets since breast cancer drug targets, approved and experimental, are found at significantly lower ranks than non-drug targets.

8.11 Prediction of target-drug and drug-drug combinations

We developed a crosstalk inhibition measure to determine the amount of crosstalk signaling that can be prevented between pathways by inhibiting specific proteins simultaneously. The concept of pathway crosstalk refers to shared protein interactions between distinct pathways. Since these interactions might also influence the downstream signaling within a pathway, the concept also comprises proteins and interactions downstream of the respective crosstalking interactions, in other words, the extended crosstalk.

Given two pathways, we first determined the potential crosstalk between them by extracting interactions, both directly or indirectly, involved in the crosstalk, and representing them as a directed network. Using this crosstalk representation, we then applied a topology-based measure, namely network efficiency (Csermely et al., 2005) to determine the signaling, that is, the flow of information, within the network. Network efficiency (*NE*) is defined as the sum of the inverse length of the shortest path between all network elements and can be computed as follows:

$$NE = \frac{\sum_{i \neq j} \frac{1}{d(i,j)}}{N(N-1)},$$

with N representing the number of network elements and d denoting the shortest distance between two elements i, j. The network efficiency ranges between 0 and 1, where 1 indicates that all proteins communicate directly with each other, i.e., a fully connected network.

Using the network efficiency determined for crosstalking pathways, as described above, we simulated the inhibition of specific protein target(s) and measured the amount of signaling that persists when removing affected proteins and interactions (NE_X). We determined the relative reduction of network efficiency, i.e., the crosstalk inhibition, as follows:

$$CI = 1 - \frac{NE_X}{NE}.$$

Since proteins are commonly annotated with more than one pathway, we computed the average *CI* across all pairs of crosstalking pathways when assessing the simultaneous inhibition of two (sets of) proteins.

To finally predict combinations, we simulated the impact of combining (1) candidate drug targets and (2) primary drug targets with breast cancer drugs having at least one protein target. When considering combinations involving novel candidate targets we were particularly interested in candidates with a high individual or complementary contribution to the crosstalk reduction. Therefore, we defined the benefit of a candidate to the crosstalk inhibition as follows:

$$B_{CT} = \frac{CI_C}{CI_{PT}}.$$

CI_{PT} and CI_C indicate the crosstalk inhibition determined when inhibiting the primary breast cancer target alone (PT) or in combination with a candidate target (C).

CHAPTER

VI

BIBLIOGRAPHY

Abbas, T., Keaton, M.A., and Dutta, A. (2013). Genomic instability in cancer. Cold Spring Harb Perspect Biol *5*, a012914.

Abd El-Rehim, D.M., Pinder, S.E., Paish, C.E., Bell, J., Blamey, R.W., Robertson, J.F., Nicholson, R.I., and Ellis, I.O. (2004). Expression of luminal and basal cytokeratins in human breast carcinoma. J Pathol *203*, 661-671.

Adam, P.J., Boyd, R., Tyson, K.L., Fletcher, G.C., Stamps, A., Hudson, L., Poyser, H.R., Redpath, N., Griffiths, M., Steers, G.*, et al.* (2003). Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J Biol Chem *278*, 6482-6489.

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.-C., De Moor, B., Marynen, P., Hassan, B.*, et al.* (2006). Gene prioritization through genomic data fusion. Nat Biotech *24*, 537-544.

Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. Nature *406*, 378-382.

Aloy, P., and Russell, R.B. (2002). The third dimension for protein interactions and complexes. Trends Biochem Sci *27*, 633-638.

Antoni, L., Sodha, N., Collins, I., and Garrett, M.D. (2007). CHK2 kinase: cancer susceptibility and cancer therapy - two sides of the same coin? Nat Rev Cancer *7*, 925-936.

Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., Ghanbarian, A.T., Kerrien, S., Khadake, J.*, et al.* (2010). The IntAct molecular interaction database in 2010. Nucleic Acids Res *38*, D525-531.

Arkin, M. (2005). Protein-protein interactions and cancer: small molecules going in for the kill. Curr Opin Chem Biol *9*, 317-324.

Bader, J.S., Chaudhuri, A., Rothberg, J.M., and Chant, J. (2004). Gaining confidence in high-throughput protein interaction networks. Nat Biotechnol *22*, 78-85.

Badie, S., Liao, C., Thanasoula, M., Barber, P., Hill, M.A., and Tarsounas, M. (2009). RAD51C facilitates checkpoint signaling by promoting CHK2 phosphorylation. J Cell Biol *185*, 587-600.

Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., and Wooster, R. (2004). The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer *91*, 355-358.

Barabasi, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat Rev Genet *12*, 56-68.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet *5*, 101-113.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D.*, et al.* (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607.

Baum, M., Brinkley, D.M., Dossett, J.A., McPherson, K., Patterson, J.S., Rubens, R.D., Smiddy, F.G., Stoll, B.A., Wilson, A., Lea, J.C.*, et al.* (1983). Improved survival among patients treated with adjuvant tamoxifen after mastectomy for early breast cancer. Lancet *2*, 450.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R.*, et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53-59.

Beraud-Dufour, S., Gautier, R., Albiges-Rizo, C., Chardin, P., and Faurobert, E. (2007). Krit 1 interactions with microtubules and membranes are regulated by Rap1 and integrin cytoplasmic domain associated protein-1. FEBS J *274*, 5518-5532.

Bernards, R. (2012). A missing link in genotype-directed cancer therapy. Cell *151*, 465-468.

Bonner, W.M., Redon, C.E., Dickey, J.S., Nakamura, A.J., Sedelnikova, O.A., Solier, S., and Pommier, Y. (2008). GammaH2AX and cancer. Nat Rev Cancer *8*, 957-967.

Bouwman, P., and Jonkers, J. (2012). The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. Nat Rev Cancer *12*, 587-598.

Breitkreutz, B.J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D.H., Bahler, J., Wood, V.*, et al.* (2008). The BioGRID Interaction Database: 2008 update. Nucleic Acids Res *36*, D637-640.

Brenton, J.D., Carey, L.A., Ahmed, A.A., and Caldas, C. (2005). Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol *23*, 7350-7360.

Bruckner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci *10*, 2763-2788.

Byrnes, G.B., Southey, M.C., and Hopper, J.L. (2008). Are the so-called low penetrance breast cancer genes, ATM, BRIP1, PALB2 and CHEK2, high risk for women with strong family histories? Breast Cancer Res *10*, 208.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C.*, et al.* (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet *40*, 722-729.

Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., Castagnoli, L., and Cesareni, G. (2010). MINT, the molecular interaction database: 2009 update. Nucleic Acids Res *38*, D532-539.

Cleaver, J.E., Lam, E.T., and Revet, I. (2009). Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. Nat Rev Genet *10*, 756-768.

Consortium, T.G.O. (2010a). The Gene Ontology in 2010: extensions and refinements. Nucleic Acids Res *38*, D331-335.

Consortium, T.U. (2010b). The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res *38*, D142-148.

Couto, F.M., Silva, M.r.J., and Coutinho, P.M. (2007). Measuring semantic similarity between Gene Ontology terms. Data & Knowledge Engineering *61*, 137-152.

Csermely, P., Agoston, V., and Pongor, S. (2005). The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci *26*, 178-182.

Chao, C., Herr, D., Chun, J., and Xu, Y. (2006). Ser18 and 23 phosphorylation is required for p53-dependent apoptosis and tumor suppression. EMBO J *25*, 2615-2622.

Chatr-Aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L.*, et al.* (2012). The BioGRID interaction database: 2013 update. Nucleic Acids Res *41*, D816-823.

Chen, F., Nastasi, A., Shen, Z., Brenneman, M., Crissman, H., and Chen, D.J. (1997). Cell cycle-dependent protein expression of mammalian homologs of yeast DNA double-strand break repair genes Rad51 and Rad52. Mutat Res *384*, 205-211.

Chen, S., and Parmigiani, G. (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol *25*, 1329-1333.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T.*, et al.* (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell *10*, 529-541.

Chou, T.C. (2006). Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. Pharmacol Rev *58*, 621-681.

Chou, T.C., and Talalay, P. (1984). Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Adv Enzyme Regul *22*, 27-55.

Dancey, J.E., and Chen, H.X. (2006). Strategies for optimizing combinations of molecularly targeted anticancer agents. Nat Rev Drug Discov *5*, 649-659.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C.*, et al.* (2002). A genomic regulatory network for development. Science *295*, 1669-1678.

Davies, A.A., Masson, J.Y., McIlwraith, M.J., Stasiak, A.Z., Stasiak, A., Venkitaraman, A.R., and West, S.C. (2001). Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. Mol Cell *7*, 273-282.

De Las Rivas, J., and Fontanillo, C. (2010). Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Biol *6*, e1000807.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) *39*, 1-38.

DeVita, V.T., Jr., Young, R.C., and Canellos, G.P. (1975). Combination versus single agent chemotherapy: a review of the basis for selection of drug treatment of cancer. Cancer *35*, 98-110.

Dinkelmann, M., Spehalski, E., Stoneham, T., Buis, J., Wu, Y., Sekiguchi, J.M., and Ferguson, D.O. (2009). Multiple functions of MRN in end-joining pathways during isotype class switching. Nat Struct Mol Biol *16*, 808-813.

Dittrich, M.T., Klau, G.W., Rosenwald, A., Dandekar, T., and Muller, T. (2008). Identifying functional modules in protein-protein interaction networks: an integrated exact approach. Bioinformatics *24*, i223-231.

Du, L., Subauste, M.C., DeSevo, C., Zhao, Z., Baker, M., Borkowski, R., Schageman, J.J., Greer, R., Yang, C.-R., Suraokar, M.*, et al.* (2012). miR-337-3p and Its Targets <italic>STAT3</italic> and <italic>RAP1A</italic> Modulate Taxane Sensitivity in Non-Small Cell Lung Cancers. PLoS One *7*, e39167.

Du, Z., Li, L., Chen, C.F., Yu, P.S., and Wang, J.Z. (2009). G-SESAME: web tools for GO-term-based gene similarity analysis and knowledge discovery. Nucleic Acids Res *37*, W345-349.

Dumitrescu, R.G., and Cotarla, I. (2005). Understanding breast cancer risk -- where do we stand in 2005? J Cell Mol Med *9*, 208-221.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., Van Tine, B.A., Hoog, J., Goiffon, R.J., Goldstein, T.C.*, et al.* (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature *486*, 353-360.

Emig, D., Ivliev, A., Pustovalova, O., Lancashire, L., Bureeva, S., Nikolsky, Y., and Bessarabova, M. (2013). Drug target prediction and repositioning using an integrated network-based approach. PLoS One *8*, e60618.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res *30*, 1575-1584.

Erten, S., Bebek, G., and Koyuturk, M. (2011). Vavien: an algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks. J Comput Biol *18*, 1561-1574.

Fearon, E.R. (2011). Molecular genetics of colorectal cancer. Annu Rev Pathol *6*, 479-507.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759-767.

Fedele, P., Calvani, N., Marino, A., Orlando, L., Schiavone, P., Quaranta, A., and Cinieri, S. (2012). Targeted agents to reverse resistance to endocrine therapy in metastatic breast cancer: where are we now and where are we going? Crit Rev Oncol Hematol *84*, 243-251.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide. In IARC CancerBase No. 11 [Internet].

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K.*, et al.* (2010). The Pfam protein families database. Nucleic Acids Res *38*, D211-222.

Forbes, S.A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., Kok, C.Y., Jia, M., Ewing, R., Menzies, A.*, et al.* (2010). COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res *38*, D652-657.

Forget, A.L., Bennett, B.T., and Knight, K.L. (2004). Xrcc3 is recruited to DNA double strand breaks early and independent of Rad51. J Cell Biochem *93*, 429-436.

Fujita, N., Jaye, D.L., Kajita, M., Geigerman, C., Moreno, C.S., and Wade, P.A. (2003). MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. Cell *113*, 207-219.

Gandhi, T.K., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K.N., Mohan, S.S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B.*, et al.* (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet *38*, 285-293.

Garattini, S. (2007). Pharmacokinetics in cancer chemotherapy. Eur J Cancer *43*, 271-282.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J.*, et al.* (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570-575.

Gietz, R.D., and Schiestl, R.H. (2007). Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc *2*, 38-41.

Gildemeister, O.S., Sage, J.M., and Knight, K.L. (2009). Cellular redistribution of Rad51 in response to DNA damage: novel role for Rad51C. J Biol Chem *284*, 31945-31952.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E.*, et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabasi, A.L. (2007). The human disease network. Proc Natl Acad Sci U S A *104*, 8685-8690.

Gottesman, M.M. (2002). Mechanisms of cancer drug resistance. Annu Rev Med *53*, 615-627.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C.*, et al.* (2007). Patterns of somatic mutation in human cancer genomes. Nature *446*, 153-158.

Greshock, J., Bachman, K.E., Degenhardt, Y.Y., Jing, J., Wen, Y.H., Eastman, S., McNeil, E., Moy, C., Wegrzyn, R., Auger, K.*, et al.* (2010). Molecular target class is predictive of in vitro response profile. Cancer Res *70*, 3677-3686.

Guimaraes, K.S., Jothi, R., Zotenko, E., and Przytycka, T.M. (2006). Predicting domain-domain interactions using a parsimony approach. Genome Biol *7*, R104.

Gundem, G., Perez-Llamas, C., Jene-Sanz, A., Kedzierska, A., Islam, A., Deu-Pons, J., Furney, S.J., and Lopez-Bigas, N. (2010). IntOGen: integration and data mining of multidimensional oncogenomic data. Nat Methods *7*, 92-93.

Guney, E., Sanz-Pamplona, R., Sierra, A., and Oliva, B. (2012). Understanding Cancer Progression Using Protein Interaction Networks. In Systems Biology in Cancer Research and Drug Discovery (Springer Netherlands), pp. 167-195.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hartlerode, A.J., and Scully, R. (2009). Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J *423*, 157-168.

Hauser, R., Ceol, A., Rajagopala, S.V., Mosca, R., Siszler, G., Wermke, N., Sikorski, P., Schwarz, F., Schick, M., Wuchty, S.*, et al.* (2014). A Second-generation Protein-Protein Interaction Network of Helicobacter pylori. Mol Cell Proteomics.

Higgins, M.J., and Baselga, J. (2011). Targeted therapies for breast cancer. J Clin Invest *121*, 3797-3803.

Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. Nature *411*, 366-374.

Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. N Engl J Med *361*, 1475-1485.

Hood, L., and Perlmutter, R.M. (2004). The impact of systems approaches on biological problems in drug discovery. Nat Biotechnol *22*, 1215-1217.

Howell, A., Cuzick, J., Baum, M., Buzdar, A., Dowsett, M., Forbes, J.F., Hoctin-Boes, G., Houghton, J., Locker, G.Y., and Tobias, J.S. (2005). Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. Lancet *365*, 60-62.

Hsieh, T.C., Wijeratne, E.K., Liang, J.Y., Gunatilaka, A.L., and Wu, J.M. (2005). Differential control of growth, cell cycle progression, and expression of NF-kappaB in human breast cancer cells MCF-7, MCF-10A, and MDA-MB-231 by ponicidin and oridonin, diterpenoids from the chinese herb Rabdosia rubescens. Biochem Biophys Res Commun *337*, 224-231.

Hudis, C.A. (2007). Trastuzumab - Mechanism of Action and Use in Clinical Practice. New England Journal of Medicine *357*, 39-51.

Hughes, J.P., Rees, S., Kalindjian, S.B., and Philpott, K.L. (2011). Principles of early drug discovery. Br J Pharmacol *162*, 1239-1249.

Isserlin, R., El-Badrawi, R.A., and Bader, G.D. (2011). The Biomolecular Interaction Network Database in PSI-MI 2.5. Database (Oxford) *2011*, baq037.

Jaeger, S., and Aloy, P. (2012). From protein interaction networks to novel therapeutic strategies. IUBMB Life *64*, 529-537.

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41-42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. Nature *407*, 651-654.

Jia, J., Zhu, F., Ma, X., Cao, Z., Li, Y., and Chen, Y.Z. (2009). Mechanisms of drug combinations: interaction and network perspectives. Nat Rev Drug Discov *8*, 111-128.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol *7*, 335-346.

Kajita, M., McClinic, K.N., and Wade, P.A. (2004). Aberrant expression of the transcription factors snail and slug alters the response to genotoxic stress. Mol Cell Biol *24*, 7559-7566.

Kato, Y., Tapping, R.I., Huang, S., Watson, M.H., Ulevitch, R.J., and Lee, J.D. (1998). Bmk1/Erk5 is required for cell proliferation induced by epidermal growth factor. Nature *395*, 713-716.

Kemp, Z.E., Carvajal-Carmona, L.G., Barclay, E., Gorman, M., Martin, L., Wood, W., Rowan, A., Donohue, C., Spain, S., Jaeger, E.*, et al.* (2006). Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. Cancer Res *66*, 5003-5006.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U.*, et al.* (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Res *40*, D841-846.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A.*, et al.* (2009). Human Protein Reference Database--2009 update. Nucleic Acids Res *37*, D767-772.

Kitano, H. (2002). Computational systems biology. Nature *420*, 206-210.

Kitano, H. (2004). Biological robustness. Nat Rev Genet *5*, 826-837.

Koegl, M., and Uetz, P. (2008). Improving yeast two-hybrid screening systems. Brief Funct Genomic Proteomic *6*, 302-312.

Kucherlapati, R., and Wheeler, D. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330-337.

Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P.*, et al.* (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics *89*, 307-315.

Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell *149*, 780-794.

Lehner, B., and Fraser, A.G. (2004). A first-draft human protein-interaction map. Genome Biol *5*, R63.

Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. Nucleic Acids Res *37*, D229-232.

Li, W., Hesabi, B., Babbo, A., Pacione, C., Liu, J., Chen, D.J., Nickoloff, J.A., and Shen, Z. (2000). Regulation of double-strand break-induced mammalian homologous recombination by UBL1, a RAD51-interacting protein. Nucleic Acids Res *28*, 1145-1153.

Liang, F., Matrubutham, U., Parvizi, B., Yen, J., Duan, D., Mirchandani, J., Hashima, S., Nguyen, U., Ubil, E., Loewenheim, J.*, et al.* (2004). ORFDB: an information resource linking scientific content to a high-quality Open Reading Frame (ORF) collection. Nucleic Acids Res *32*, D595-599.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E.*, et al.* (2012). MINT, the molecular interaction database: 2012 update. Nucleic Acids Res *40*, D857-861.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jorgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K.*, et al.* (2007). Systematic discovery of in vivo phosphorylation networks. Cell *129*, 1415-1426.

Liu, Y., Hu, B., Fu, C., and Chen, X. (2010). DCDB: drug combination database. Bioinformatics *26*, 587-588.

Liu, Y., Tarsounas, M., O'Regan, P., and West, S.C. (2007). Role of RAD51C and XRCC3 in genetic recombination and DNA repair. J Biol Chem *282*, 1973-1979.

Lord, C.J., and Ashworth, A. (2012). The DNA damage response and cancer therapy. Nature *481*, 287-294.

Luo, J., Emanuele, M.J., Li, D., Creighton, C.J., Schlabach, M.R., Westbrook, T.F., Wong, K.K., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell *137*, 835-848.

Markowitz, S.D., and Bertagnolli, M.M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med *361*, 2449-2460.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B.*, et al.* (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res *37*, D619-622.

McKusick, V.A. (2007). Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet *80*, 588-604.

Mewes, H.W., Ruepp, A., Theis, F., Rattei, T., Walter, M., Frishman, D., Suhre, K., Spannagl, M., Mayer, K.F., Stumpflen, V., and Antonov, A. (2011). MIPS: curated databases and comprehensive secondary data resources in 2010. Nucleic Acids Res *39*, D220-224.

Miller, T.W. (2013). Endocrine resistance. Am Soc Clin Oncol Educ Book, 37-42.

Mladenov, E., and Iliakis, G. (2011). The Pathways of Double-Strand Break Repair.

Mosca, R., Céol, A., and Aloy, P. (2013). Interactome3D: adding structural details to protein networks. Nature methods *10*, 47-53.

Moyal, L., Lerenthal, Y., Gana-Weisz, M., Mass, G., So, S., Wang, S.Y., Eppink, B., Chung, Y.M., Shalev, G., Shema, E.*, et al.* (2011). Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. Mol Cell *41*, 529-542.

Nakamura, K., Kato, A., Kobayashi, J., Yanagihara, H., Sakamoto, S., Oliveira, D.V., Shimada, M., Tauchi, H., Suzuki, H., Tashiro, S.*, et al.* (2011). Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. Mol Cell *41*, 515-528.

Navlakha, S., and Kingsford, C. (2010). The power of protein interaction networks for associating genes with diseases. Bioinformatics *26*, 1057-1063.

Neumann, B., Held, M., Liebel, U., Erfle, H., Rogers, P., Pepperkok, R., and Ellenberg, J. (2006). High-throughput RNAi screening by time-lapse imaging of live human cells. Nat Methods *3*, 385-390.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F.*, et al.* (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515-527.

Nielsen, T.O., Hsu, F.D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L.*, et al.* (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res *10*, 5367-5374.

Obayashi, T., Okamura, Y., Ito, S., Tadaka, S., Motoike, I.N., and Kinoshita, K. (2013). COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals. Nucleic Acids Res *41*, D1014-1020.

Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., Bridge, A., Briganti, L., Brinkman, F.S., Cesareni, G.*, et al.* (2012). Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat Methods *9*, 345-350.

Osborne, C.K., and Schiff, R. (2011). Mechanisms of endocrine resistance in breast cancer. Annu Rev Med *62*, 233-247.

Osborne, C.K., Shou, J., Massarweh, S., and Schiff, R. (2005). Crosstalk between estrogen receptor and growth factor receptor pathways as a cause for endocrine therapy resistance in breast cancer. Clin Cancer Res *11*, 865s-870s.

Oti, M., and Brunner, H.G. (2007). The modular nature of genetic diseases. Clin Genet *71*, 1-11.

Oti, M., Snel, B., Huynen, M.A., and Brunner, H.G. (2006). Predicting disease genes using protein-protein interactions. J Med Genet *43*, 691-698.

Pammolli, F., Magazzini, L., and Riccaboni, M. (2011). The productivity crisis in pharmaceutical R&D. Nat Rev Drug Discov *10*, 428-438.

Parsons, J.L., and Dianov, G.L. (2013). Co-ordination of base excision repair and genome stability. DNA Repair (Amst) *12*, 326-333.

Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., and Schacht, A.L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov *9*, 203-214.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A.*, et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Polyak, K. (2007). Breast cancer: origins and evolution. J Clin Invest *117*, 3155-3163.

Polyak, K. (2011). Heterogeneity in breast cancer. J Clin Invest *121*, 3786-3788.

Polyak, K., and Metzger Filho, O. (2012). SnapShot: breast cancer. Cancer Cell *22*, 562-562 e561.

Prasad, T.S., Kandasamy, K., and Pandey, A. (2009). Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. Methods Mol Biol *577*, 67-79.

Prat, A., and Perou, C.M. (2011). Deconstructing the molecular portraits of breast cancer. Mol Oncol *5*, 5-23.

Pujol, A., Mosca, R., Farres, J., and Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. Trends Pharmacol Sci *31*, 115-123.

Raguz, S., and Yague, E. (2008). Resistance to chemotherapy: new treatments and novel insights into an old problem. Br J Cancer *99*, 387-391.

Rajagopala, S.V., Hughes, K.T., and Uetz, P. (2009). Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins. Proteomics *9*, 5296-5302.

Rajagopala, S.V., Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., Franca-Koh, J., Pakala, S.B., Phanse, S., Ceol, A.*, et al.* (2014). The binary protein-protein interaction landscape of Escherichia coli. Nat Biotechnol *32*, 285-290.

Rajagopalan, H., Nowak, M.A., Vogelstein, B., and Lengauer, C. (2003). The significance of unstable chromosomes in colorectal cancer. Nat Rev Cancer *3*, 695-701.

Ramaswamy, S. (2007). Rational design of cancer-drug combinations. N Engl J Med *357*, 299-300.

Reinhardt, H.C., and Schumacher, B. (2012). The p53 network: cellular and systemic DNA damage responses in aging and cancer. Trends Genet *28*, 128-136.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., and Briggs, B.B. (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia *9*, 166-180.

Rhodes, D.R., Tomlins, S.A., Varambally, S., Mahavisno, V., Barrette, T., Kalyana-Sundaram, S., Ghosh, D., Pandey, A., and Chinnaiyan, A.M. (2005). Probabilistic model of the human protein-protein interaction network. Nat Biotechnol *23*, 951-959.

Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A.M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*, 1-6.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol *17*, 1030-1032.

Rouleau, M., Patel, A., Hendzel, M.J., Kaufmann, S.H., and Poirier, G.G. (2010). PARP inhibition: PARP1 and beyond. Nat Rev Cancer *10*, 293-301.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N.*, et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Russell, R.B., and Aloy, P. (2008). Targeting and tinkering with interaction networks. Nat Chem Biol *4*, 666-673.

Sabates-Bellver, J., Van der Flier, L.G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M.A., Bujnicki, J.M., Menigatti, M.*, et al.* (2007). Transcriptome profile of human colorectal adenomas. Molecular cancer research : MCR *5*, 1263-1275.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. Nucleic Acids Res *32*, D449-451.

Sambrano, G.R., Chandy, G., Choi, S., Decamp, D., Hsueh, R., Lin, K.M., Mock, D., O'Rourke, N., Roach, T., Shu, H.*, et al.* (2002). Unravelling the signal-transduction network in B lymphocytes. Nature *420*, 708-710.

Sancho, E., Batlle, E., and Clevers, H. (2004). Signaling pathways in intestinal development and cancer. Annu Rev Cell Dev Biol *20*, 695-723.

Sawyers, C. (2004). Targeted cancer therapy. Nature *432*, 294-297.

Shah, S., and Federoff, H.J. (2009). Drug discovery dilemma and Cura quartet collaboration. Drug Discov Today *14*, 1006-1010.

Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. Mol Syst Biol *3*, 88.

Shema, E., Tirosh, I., Aylon, Y., Huang, J., Ye, C., Moskovits, N., Raver-Shapira, N., Minsky, N., Pirngruber, J., Tarcic, G.*, et al.* (2008). The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. Genes & development *22*, 2664-2676.

Shen, Z. (2012). Genomic instability and cancer: an introduction. J Mol Cell Biol *3*, 1-3.

Shiloh, Y., Shema, E., Moyal, L., and Oren, M. (2011). RNF20-RNF40: A ubiquitin-driven link between gene expression and the DNA damage response. FEBS Lett *585*, 2795-2802.

Shoemaker, B.A., and Panchenko, A.R. (2007). Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol *3*, e42.

Shoemaker, R.H. (2006). The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer *6*, 813-823.

Shou, J., Massarweh, S., Osborne, C.K., Wakeling, A.E., Ali, S., Weiss, H., and Schiff, R. (2004). Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. J Natl Cancer Inst *96*, 926-935.

Simonis, N., Rual, J.F., Carvunis, A.R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Sahalie, J.M., Venkatesan, K., Gebreab, F.*, et al.* (2009). Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. Nat Methods *6*, 47-54.

Sjoblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N.*, et al.* (2006). The consensus coding sequences of human breast and colorectal cancers. Science *314*, 268-274.

Skoglund, J., Djureinovic, T., Zhou, X.L., Vandrovcova, J., Renkonen, E., Iselius, L., Bisgaard, M.L., Peltomaki, P., and Lindblom, A. (2006). Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. J Med Genet *43*, e7.

Skrabanek, L., and Campagne, F. (2001). TissueInfo: high-throughput identification of tissue expression profiles and specificity. Nucleic Acids Res *29*, E102-102.

Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M.*, et al.* (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med *344*, 783-792.

Smith, C. (2003). Drug target validation: Hitting the target. Nature *422*, 341, 343, 345 passim.

Smith, D.C., Smith, M.R., Sweeney, C., Elfiky, A.A., Logothetis, C., Corn, P.G., Vogelzang, N.J., Small, E.J., Harzstark, A.L., Gordon, M.S.*, et al.* (2013). Cabozantinib in patients with advanced prostate cancer: results of a phase II randomized discontinuation trial. J Clin Oncol *31*, 412-419.

Spirin, V., and Mirny, L.A. (2003). Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A *100*, 12123-12128.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S.*, et al.* (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell *122*, 957-968.

Stracker, T.H., and Petrini, J.H. (2011). The MRE11 complex: starting from the ends. Nat Rev Mol Cell Biol *12*, 90-103.

Stracker, T.H., Usui, T., and Petrini, J.H. (2009). Taking the time to make important decisions: the checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. DNA Repair (Amst) *8*, 1047-1054.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719-724.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249-255.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G.*, et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A *101*, 6062-6067.

Su, Z.Z., Lebedeva, I.V., Sarkar, D., Gopalkrishnan, R.V., Sauane, M., Sigmon, C., Yacoub, A., Valerie, K., Dent, P., and Fisher, P.B. (2003). Melanoma differentiation associated gene-7, mda-7/IL-24, selectively induces growth suppression, apoptosis and radiosensitization in malignant gliomas in a p53-independent manner. Oncogene *22*, 1164-1180.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Sy, S.M., Huen, M.S., and Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc Natl Acad Sci U S A *106*, 7155-7160.

Thompson, D., and Easton, D. (2004). The genetic epidemiology of breast cancer genes. J Mammary Gland Biol Neoplasia *9*, 221-236.

Tsang, R.Y., and Finn, R.S. (2012). Beyond trastuzumab: novel therapeutic strategies in HER2-positive metastatic breast cancer. Br J Cancer *106*, 6-13.

Turnbull, C., and Rahman, N. (2008). Genetic predisposition to breast cancer: past, present, and future. Annu Rev Genomics Hum Genet *9*, 321-345.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.*, et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C.*, et al.* (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics *4*, 1920-1932.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T.*, et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

van der Greef, J., and McBurney, R.N. (2005). Innovation: Rescuing drug discovery: in vivo systems pathology and systems pharmacology. Nat Rev Drug Discov *4*, 961-967.

Vazquez, A., Flammini, A., Maritan, A., and Vespignani, A. (2003). Global protein function prediction from protein-protein interaction networks. Nat Biotechnol *21*, 697-700.

Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I.*, et al.* (2009). An empirical framework for binary interactome mapping. Nat Methods *6*, 83-90.

Vlasblom, J., and Wodak, S.J. (2009). Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. BMC Bioinformatics *10*, 99.

Vogel, C.L., Cobleigh, M.A., Tripathy, D., Gutheil, J.C., Harris, L.N., Fehrenbacher, L., Slamon, D.J., Murphy, M., Novotny, W.F., Burchmore, M.*, et al.* (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol *20*, 719-726.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417*, 399-403.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer *9*, 489-499.

Wang, C., Deng, L., Hong, M., Akkaraju, G.R., Inoue, J., and Chen, Z.J. (2001). TAK1 is a ubiquitin-dependent kinase of MKK and IKK. Nature *412*, 346-351.

Wang, J., Li, M., Deng, Y., and Pan, Y. (2010). Recent advances in clustering methods for protein interaction networks. BMC Genomics *11 Suppl 3*, S10.

Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S., and Chen, C.F. (2007). A new method to measure the semantic similarity of GO terms. Bioinformatics *23*, 1274-1281.

Weigelt, B., and Reis-Filho, J.S. (2009). Histological and molecular types of breast cancer: is there a unifying taxonomy? Nat Rev Clin Oncol *6*, 718-730.

Wiesner, G.L., Daley, D., Lewis, S., Ticknor, C., Platzer, P., Lutterbaugh, J., MacMillen, M., Baliner, B., Willis, J., Elston, R.C., and Markowitz, S.D. (2003). A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci U S A *100*, 12961-12965.

Williams, C.D., Satia, J.A., Adair, L.S., Stevens, J., Galanko, J., Keku, T.O., and Sandler, R.S. (2011). Associations of red meat, fat, and protein intake with distal colorectal cancer risk. Nutr Cancer *62*, 701-709.

Wong, Y.H., Lee, T.Y., Liang, H.K., Huang, C.M., Wang, T.Y., Yang, Y.H., Chu, C.H., Huang, H.D., Ko, M.T., and Hwang, J.K. (2007). KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. Nucleic Acids Res *35*, W588-594.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J.*, et al.* (2007). The genomic landscapes of human breast and colorectal cancers. Science *318*, 1108-1113.

Wood, R.D., Mitchell, M., and Lindahl, T. (2005). Human DNA repair genes, 2005. Mutat Res *577*, 275-283.

Yacoub, A., Mitchell, C., Lebedeva, I.V., Sarkar, D., Su, Z.Z., McKinstry, R., Gopalkrishnan, R.V., Grant, S., Fisher, P.B., and Dent, P. (2003). mda-7 (IL-24) Inhibits growth and enhances radiosensitivity of glioma cells in vitro via JNK signaling. Cancer Biol Ther *2*, 347-353.

Yamaguchi, H., Chang, S.S., Hsu, J.L., and Hung, M.C. (2013a). Signaling cross-talk in the resistance to HER family receptor targeted therapy. Oncogene.

Yamaguchi, H., Chang, S.S., Hsu, J.L., and Hung, M.C. (2013b). Signaling cross-talk in the resistance to HER family receptor targeted therapy. Oncogene *33*, 1073-1081.

Yang, Q., Deng, X., Lu, B., Cameron, M., Fearns, C., Patricelli, M.P., Yates, J.R., III, Gray, N.S., and Lee, J.-D. (2010). Pharmacological Inhibition of BMK1 Suppresses Tumor Growth through Promyelocytic Leukemia Protein. Cancer Cell *18*, 258-267.

Yeh, S.H., Yeh, H.Y., and Soo, V.W. (2012). A network flow approach to predict drug targets from microarray data, disease genes and interactome network - case study on prostate cancer. J Clin Bioinforma *2*, 1.

Yook, S.H., Oltvai, Z.N., and Barabasi, A.L. (2004). Functional and topological characterization of protein interaction networks. Proteomics *4*, 928-942.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics *26*, 976-978.

Zhang, J., and Huan, J. (2010). Novel biological network features discovery for in silico identification of drug targets. In Proceedings of the 1st ACM International Health Informatics Symposium (Arlington, Virginia, USA, ACM).

Zhu, F., Shi, Z., Qin, C., Tao, L., Liu, X., Xu, F., Zhang, L., Song, Y., Zhang, J., Han, B.*, et al.* (2012). Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. Nucleic Acids Res *40*, D1128-1136.

Zhu, M., Gao, L., Li, X., Liu, Z., Xu, C., Yan, Y., Walker, E., Jiang, W., Su, B., Chen, X., and Lin, H. (2009). The analysis of the drug-targets based on the topological properties in the human protein-protein interaction network. J Drug Target *17*, 524-532.

Zou, J.X., Revenko, A.S., Li, L.B., Gemo, A.T., and Chen, H.W. (2007). ANCCA, an estrogen-regulated AAA+ ATPase coactivator for ERalpha, is required for coregulator occupancy and chromatin modification. Proc Natl Acad Sci U S A *104*, 18067-18072.

CHAPTER

VII

SUMMARY IN SPANISH

1. Introducción

Las ciencias biomédicas han estado tradicionalmente inmersas en un reduccionismo conceptual, centrándose principalmente en el estudio detallado de genes y proteínas individuales. Décadas de investigación en biología celular, molecular y estructural han aumentado considerablemente nuestra comprensión acerca de las proteínas individuales que participan en los procesos biológicos. Sin embargo, los sistemas biológicos son complejos por naturaleza, y el estudio individualizado de sus componentes revela relativamente poco acerca de su función y organización. Como las proteínas rara vez actúan solas, el enfoque tradicional es incapaz de predecir el comportamiento de un organismo intacto y la forma en que éste cambia de forma coordinada en respuesta a un estímulo particular, como la aparición de una enfermedad.

Las ciencias farmacológicas han seguido un curso similar, con enfoques tradicionales centrados en el estudio, a nivel molecular, de la dupla fármaco-diana terapéutica. Muchos fármacos con grandes expectativas han fracasado rotundamente las últimas fases clínicas porque los mecanismos de acción de las vías a las que se dirigen son todavía desconocidos (Pammolli et al., 2011). Estos efectos se han acentuado en la última década, cuando la investigación farmacéutica se ha centrado en enfermedades cada vez más complejas y menos conocidas, como el cáncer.

El cáncer de mama y el cáncer colorrectal son un perfecto ejemplo de enfermedad compleja que, a pesar de muchos años de investigación, está lejos de ser bien comprendida. Varios trabajos realizados estudiando los patrones de expresión génica, las mutaciones, el número de copias de ADN, y otras modificaciones proteicas que ocurren en los carcinomas de mama han incrementado enormemente el conocimiento de los actores involucrados en la progresión de dicha enfermedad (Barretina et al., 2012; Ellis et al., 2012; Garnett et al., 2012; Perou et al., 2000). Sin embargo, la heterogeneidad del cáncer rara vez se debe a anomalías en genes individuales, sino más bien refleja la interrupción de complejos procesos intra e intercelulares (Barabasi et al., 2011). Por tanto, la comunidad científica se está desplazando hacia enfoques de sistemas, en los cuales se examinan las propiedades globales.

Un modo útil de describir y analizar la heterogeneidad del cáncer es el uso de la biología de sistemas. La biología de redes, concretamente, es una disciplina basada en redes que estudia las interacciones entre moléculas, y su atención se centra en las redes de interacciones entre proteínas. Enfoques basados en dichas redes permiten situar les proteínas de vuelta a su contexto, teniendo en cuenta una perspectiva mucho más amplia de su entorno sin perder los detalles moleculares. Sin embargo, la disponibilidad de un mapa completo de las interacciones proteicas que pueden producirse en un organismo (interactoma) es crucial para llevar a cabo análisis de redes. Esto implica que la mayoría de los trabajos para dar a conocer las bases moleculares de las patologías deben incluir un paso inicial de descubrimiento de interacciones. Por esta razón, la identificación sistemática de las interacciones implicadas en una determinada enfermedad es crítica, y prueba de ellos son trabajos publicados en la última década centrados en la detección de interacciones proteicas (Hauser et al., 2014; Rajagopala et al., 2014; Rual et al., 2005; Simonis et al., 2009; Stelzl et al., 2005). Entre las muchas técnicas disponibles para efectuar dicha tarea, la técnica del doble híbrido en levadura es una de las tecnologías más aplicadas con éxito a gran escala (Fields and Song, 1989).

Entre las múltiples potenciales aplicaciones biológicas y clínicas de los estudios basados en redes de interacciones proteicas podemos destacar los estudios sobre el tratamiento del cáncer. A medida que la investigación en este campo avanza, se hace más y más evidente que la búsqueda de una 'bala mágica' que derrotará a todas las formas de cáncer, de igual modo que hacen los antibióticos frente a las infecciones bacterianas, nunca tendrá éxito. Para mejorar nuestra eficacia en este campo, nuevas estrategias deberían evitar el enfoque reduccionista que supone implícitamente que la destrucción de las células cancerosas se puede lograr por sólo interferir con una única proteína. Diferentes estudios y observaciones clínicas han demostrado que los sistemas celulares son redundantes y robustos (Kitano, 2004), y las células cancerosas (o sus poblaciones) pueden encontrar la manera de escapar de un único punto de bloqueo. Sin duda, el fracaso de los tratamientos debido a la resistencia a los medicamentos sigue siendo un reto importante en la mayoría de los cánceres sólidos avanzados, como el cáncer de mama (Raguz and Yague, 2008). En su conjunto, las disciplinas basadas en redes y la biología de sistemas pueden

revolucionar el estudio de enfermedades complejas, como el cáncer de mama y colorrectal. Estos enfoques podrían ayudar significativamente en el desarrollo de nuevas y más efectivas terapias.

Esta tesis es un trabajo multidisciplinario que involucra varios enfoques por lo que, para facilitar su comprensión, he dividido el contenido en dos capítulos principales. En primer lugar, se presentan y discuten los resultados obtenidos en cuanto a la caracterización molecular de los cánceres de mama y de colon. El siguiente capítulo trata los estudios relacionados con la identificación y validación inicial de nuevas dianas terapéuticas para el cáncer de mama, así como de nuevas combinaciones de fármacos.

2. Objetivos

Capítulo I: Caracterización molecular de los cánceres de mama y colon.

1. Exploración de los mecanismos que subyacen a los genes relacionados con dichas enfermedades. El principal objetivo es descubrir y caracterizar nuevas interacciones entre genes causantes y genes asociados a los cánceres de mama y colorrectal.
2. Expansión de los interactomas de los dos cánceres estudiados con el fin de identificar nuevos genes relacionados con cada enfermedad.
3. Centrándonos en el cáncer de mama, nuestro objetivo es integrar los nuevos datos de interacción obtenidos con los datos actualmente disponibles en la literatura. A continuación aplicar una estrategia basada en biología de redes con el propósito de descubrir la función de las proteínas cuyo papel en la enfermedad permanece desconocido.

Capítulo II: Identificación y validación inicial de nuevas dianas farmacológicas y nuevas combinaciones de fármacos para el tratamiento de cáncer de mama.

1. Identificar nuevas potenciales dianas farmacológicas para el cáncer de mama basándonos en las propiedades topológicas de la red y realizar una evaluación inicial *in vitro* del efecto de la inhibición de dichas dianas.
2. Validar *in vitro* una serie de combinaciones entre potenciales dianas terapéuticas y fármacos ya aprobados para el tratamiento del cáncer de mama. Estas combinaciones han sido pronosticadas en base a una medida del solapamiento de la inhibición entre diferentes rutas de señalización.
3. Predecir y validar nuevas combinaciones entre medicamentos ya aprobados para el tratamiento del cáncer de mama utilizando la misma estrategia que anteriormente.

3. Resultados y discusión

Capítulo I: Caracterización molecular de los cánceres de mama y colon.

Expansión de los interactomas del cáncer de mama y colorrectal.

En primer lugar usamos el sistema de doble híbrido de alto rendimiento en levadura para identificar nuevas interacciones entre los genes que causan los cánceres de mama o colorrectal y los genes asociados a estos tipos de cáncer pero cuyo papel en la enfermedad es desconocido

La selección de los genes causantes de cáncer fue similar en ambos estudios: elegimos genes bien establecidos y caracterizados en el desarrollo del cáncer de mama y del cáncer de colon. Sin embargo, a pesar de que el objetivo en ambos estudios es situar los genes asociados en el interactoma de cada tipo de cáncer, en cada caso seguimos una estrategia diferente para seleccionar dichos genes asociados. En el primer estudio (cáncer de mama), la priorización se realizó en base a su nivel en la red de interacciones proteicas. En este caso nuestro objetivo fue optar por aquellos genes cuya distancia en la red era mayor, y por tanto descartamos los genes cercanos a un gen causante. De esta manera, maximizamos la probabilidad de detectar nuevas interacciones. Por otro lado, en el segundo estudio (cáncer de colon) elegimos genes cuya expresión en tejidos y líneas celulares se coordina con los genes causantes de dicho cáncer CRC. Además, se incluyeron los genes localizados en la región 9q22.32-31.1, que contiene un locus de susceptibilidad (CRCS9) que participa en el desarrollo de síndromes de cáncer colorrectal hereditario. Nuestro objetivo era una vez más detectar los genes en esta región que interactúan con los genes causantes del cáncer. Estas diferencias en la selección de los genes candidatos no afectan la relevancia biológica de los datos de interacción obtenidos, como se discute más adelante. Cada interacción fue testada por duplicado y se midió la activación de tres genes indicadores diferentes (HIS, URA3 y LACZ). Con el fin de disminuir el número de falsos positivos y para mejorar la fiabilidad de los resultados, seleccionamos las interacciones detectadas que fueron capaces de activar al menos dos genes indicadores o de activar ambas réplicas biológicas. Estas interacciones altamente fiables son las que fueron posteriormente

utilizadas para los estudios de biología de redes, de modo que se reduce el número de posibles artefactos de nuestros ensayos. En resumen, identificamos:

- 491 interacciones entre los productos de 49 genes causantes y 54 genes asociados al cáncer de mama.
- 595 interacciones entre los productos de 45 genes causante y 45 genes asociados al cáncer colorrectal.

En segundo lugar, realizamos la técnica de doble híbrido en levadura usando una librería de cDNAs con el objetivo de identificar nuevas proteínas que puedan estar relacionados con el desarrollo de los dos cánceres estudiados. Una vez más, seguimos dos estrategias diferentes para seleccionar los genes testados como *anzuelos* en el ensayo. En el caso del cáncer de mama elegimos los dos más relevantes en este cáncer (ER y HER2) y nueve genes cuyas mutaciones están asociadas a su predisposición genética. En este experimento detectamos 108 interacciones altamente fiables (siguiendo el criterio anteriormente descrito) entre estos 11 genes y 99 *presas* de la biblioteca de cDNA. A continuación verificamos que 76 de las 99 proteínas identificadas no han sido previamente relacionadas con el cáncer de mama (Gundem et al., 2010), pero en cambio sí que observamos que se encuentran co-expresadas con sus interactores en tejidos humanos, lo cual valida la estrategia seguida. En cambio, en el case del cáncer colorrectal se seleccionaron seis genes como *anzuelos* en base a perfiles de expresión génica de los tejidos normal, tumor benigno y adenocarcinoma (Sabates-Bellver et al., 2007). En particular, seleccionamos seis genes altamente expresados en los tumores colorrectales en comparación tanto con la mucosa normal como con los tumores benignos. Detectamos y validamos 27 nuevas interacciones, entre los seis cebos probados y 20 nuevas proteínas. Por consiguiente, estas 20 proteínas podrían ser relevantes en el desarrollo de CRC, y creemos que deben tomarse en consideración en futuros estudios sobre el cáncer colorrectal.

El siguiente paso fue la fusión de las interacciones obtenidas en los dos tipos de ensayos de doble híbrido efectuados, y obtuvimos dos conjuntos de 599 y 622 nuevas interacciones que definen nuestra contribución a la ampliación de los interactomas de cáncer de mama y de colon, respectivamente. A continuación llevamos a cabo dos análisis distintos para analizar e intentar fortalecer la calidad de los datos obtenidos. En primer lugar, validamos experimentalmente

un subconjunto aleatorio de interacciones utilizando técnicas complementarias, obteniendo un 79% de validación mediante co-immunoprecipitación, resultados que apoyan firmemente la confianza de nuestras interacciones. En segundo lugar, se evaluó computacionalmente si las interacciones descubiertas podrían realmente dar detalles mecanísticos sobre las relaciones entre los genes causantes de cada cáncer estudiado y los candidatos. Con dicho fin, verificamos si los genes causantes y sus interactores detectados participaban en procesos biológicos similares, en un procedimiento que hemos denominado como 'coherencia funcional' de nuestras redes de interacción. La alta coherencia observada ambos cánceres estudiados indica que nuestra estrategia basada en la biología de redes es de hecho una herramienta de inferencia robusta que permite conocer mejor los mecanismos subyacentes de esas proteínas con funciones hasta ahora desconocidas. De igual, también indica que dicha estrategia ofrece una mejor comprensión de la regulación entre diferentes proteínas presentes en los complejos sistemas biológicos.

Análisis del interactoma asociado al cáncer de mama

Los siguientes estudios fueron realizados centrándonos únicamente en el cáncer de mama. Con el fin de evaluar en profundidad la repercusión de estas nuevas interacciones identificadas, es necesario estudiarlas en su contexto biológico. Por lo tanto, integramos nuestro conjunto de 599 nuevas interacciones con los datos de interacción humana disponibles en la actualidad, obteniendo de esta manera una red de 11.226 interacciones entre 2.019 proteínas, designada como BC-PIN. A continuación analizamos la estructura de BC-PIN para detectar la presencia de módulos funcionales (grupos de proteínas densamente interconectadas y funcionalmente homogéneas). Usando el algoritmo de agrupamiento MCL (Enright et al., 2002) identificamos 178 módulos en el BC-PIN, 146 de los cuales mostraron un alto grado de homogeneidad funcional. Al volver a analizar estos módulos funcionales excluyendo nuestras 599 interacciones de HC, observamos un menor número de módulos, así como considerablemente menos módulos homogéneos. Este resultado apoya la idea de que nuestro estudio saca a relucir algunas áreas inexploradas del interactoma asociado al cáncer de mama. Posteriormente agrupamos los módulos con las mismas funciones celulares homogéneas, creando así subredes dentro de BC-PIN. Además de a la función de reparación

de ADN (procedimiento descrito a continuación), también fuimos capaces de asociar algunos de nuestros genes candidatos a una variedad de procesos tales como la apoptosis, la transducción de señales, la regulación de la transcripción, el procesamiento del ARN mensajero, el plegamiento de proteínas, o el transporte de iones, entre otros.

En resumen, la creación y posterior estudio de la red BC-PIN nos permite interpretar y medir la relevancia de nuestras nuevas interacciones, teniendo en cuenta su contexto y su peso en el interactoma del cáncer de mama.

Identificación de nuevos genes relacionados con la reparación del ADN

A continuación fusionamos todos los módulos homogéneos para la función de reparación del ADN y así construir una subred relacionada con dicha función. Esta subred contiene cuatro módulos, 87 proteínas y 167 interacciones. Posteriormente nos centramos en 15 genes asociados al cáncer de mama y provenientes de los ensayos previos de interactómica. Curiosamente, sólo uno (RNF20) de estos genes había sido previamente descrito como un modulador de la respuesta al daño de ADN, pero en cambio los 14 genes restantes nunca habían sido relacionados con este proceso. Nuestro principal objetivo en esta parte del proyecto fue la de comprobar si estos genes están realmente asociados con la respuesta al daño del ADN. Por consiguiente, sobreexpresamos y/o inhibimos estos genes y realizamos dos ensayos diferentes (el ensayo clonogénico y ensayos de formación de focos después de irradiación). En general, pudimos detectas una posible asociación de seis genes (RNF20, SERPINB5, SNAIL1, FAM84B, MTA3 e IL24) a la respuesta al daño del ADN. Estos resultados indican que, aunque se requieren más experimentos para demostrar que dichos genes juegan un papel fundamental en la reparación del ADN, el método basado en el análisis mediante módulos de la red es una estrategia fiable para la predicción nuevas funciones proteicas.

Por último, estudiamos en mayor profundidad dos genes que muestran consistentemente resultados positivos en los experimentos realizados: RNF20 y FAM84B. En cuanto a RNF20, estudios recientes han demostrado un papel en el reclutamiento de las proteínas de reparación sobre la fosforilación de ATM. Sin embargo, proponemos un papel adicional en base a dos nuevas interacciones con XRCC3 y CHK2 reportadas en nuestros ensayos de doble

híbrido en levadura. Por lo que a FAM84B se refiere, una proteína cuya función es desconocida, nuestros hallazgos sugieren que la sobreexpresión de FAM84B observada en células de mama podría promover la tumorigénesis temprana mediante la alteración de los mecanismos de reparación del ADN a través de la estabilización de RAD51.

Capítulo II: Identificación y validación inicial de nuevas dianas farmacológicas y nuevas combinaciones de fármacos para el tratamiento de cáncer de mama.

En el segundo estudio se tuvo en cuenta la redundancia y el solapamiento de vías o rutas de señalización para predecir nuevas dianas de fármacos y nuevas combinaciones de fármacos para el cáncer de mama.

Identificación de nuevas dianas terapéuticas

Partiendo de la lista de 59 genes descritos como causantes del cáncer de mama recopilados en el primer capítulo de la tesis, se generó una red de interacciones proteicas específica del cáncer de mama. Posteriormente se identificaron las proteínas de la red que son dianas terapéuticas de medicamentos usados actualmente contra el cáncer de mama. El siguiente paso fue seleccionar las proteínas candidatas a ser nuevas dianas terapéuticas: seleccionamos las proteínas implicadas en procesos biológicos similares a las dianas ya conocidas (210 candidatos iniciales), y a continuación clasificamos estos candidatos según su similitud topológica a dichas dianas terapéuticas. Nuestro objetivo en este estudio es distinguir los candidatos que se asemejan no sólo funcionalmente, sino también topológicamente a las dianas terapéuticas ya explotadas. Finalmente, acabamos con una lista definitiva de 54 nuevas posibles dianas terapéuticas.

Antes de realizar una validación experimental tradicional, se analizaron los perfiles de actividad de los moduladores químicos asociados con nuestros 54 candidatos y que habían sido previamente testados en el panel de la líneas celulares NCI60 (Shoemaker, 2006). Sorprendentemente, al inhibir los nuevos candidatos propuestos se observó una mayor actividad contra líneas celulares

del cáncer de mama en comparación, ya sea a un grupo aleatorio de todos los compuestos del NCI60, y también a un conjunto aleatorio de proteínas de la red asociada a dicho cáncer. Por otra parte, la actividad de las dianas predichas fue similar a la actividad de los fármacos contra el cáncer de mama ya aprobados. Por tanto, la evaluación de los perfiles de actividad NCI60 enfatiza la posible relevancia de las dianas de fármacos predichas.

A continuación, testamos experimentalmente mediante ensayos de MTT un grupo de ocho candidatos a diana. El experimento se realizó en cuatro líneas celulares que representan los principales subtipos de cáncer de mama, junto con una línea celular de osteosarcoma y en una línea de células no tumorigénicas. En resumen, observamos que dos candidatos a diana no muestran ningún efecto (IL1R1 y PTPN6), y la inhibición de MAP2K2 afectó principalmente a la línea de células no tumorigénicas. En cuanto a la diana PPP2R5A, aunque se llegó a alcanzar el 80% de muerte celular (IC_{80}) en cuatro líneas celulares, las dosis requeridas son sumamente altas, y el efecto más relevante se observó en la línea celular de osteosarcoma. Resultó interesante observar como la inhibición de las cuatro dianas restantes condujo a una importante inhibición del crecimiento de varias líneas celulares de cáncer de mama: mientras que la inhibición de PIK3CB fue capaz de afectar la supervivencia celular al 50-60% a lo sumo, el efecto de los fármacos dirigidos a MAP3K7, MAPK7 y RAP1A fue mucho mayor (80%). El bloqueo de MAP3K7, por ejemplo, redujo consistentemente el crecimiento celular al 80% en todas las líneas celulares (incluyendo la línea celular MCF-10A, no tumorigenica). RAP1A, por el contrario, fue capaz de afectar altamente todas las células cancerosas, pero no tuvo ningún efecto en las células MCF-10A. En conclusión, la inhibición de cuatro de los ocho objetivos testados produjo una disminución clara y consistente de la supervivencia de las células del cáncer de mama, mientras que otros dos inhibieron ligeramente el crecimiento de algunos subtipos celulares.

Identificación de nuevas combinaciones de fármacos

A continuación, quisimos evaluar si el cálculo de la inhibición de la comunicación y/o solapamiento entre distintas rutas de señalización puede ser útil para inferir nuevas combinaciones de fármacos para el tratamiento del

cáncer de mama. Por esta razón, desarrollamos una estrategia computacional para inferir pares de combinaciones de fármacos aprobados basándonos en la redundancia funcional y la interferencia entre vías de estos fármacos. De este modo, obtuvimos 413 nuevas combinaciones que, o bien superan el umbral de inhibición establecido para las combinaciones actualmente en uso o bien muestran un posible comportamiento sinérgico/aditivo. El siguiente paso de nuestra estrategia fue combinar los dos enfoques anteriores: identificación de dianas terapéuticas y predicción de nuevas combinaciones. Por lo tanto, explotado el mismo concepto anteriormente descrito simulamos el impacto de inhibir nuevos candidatos diana en combinación con fármacos en uso para tratar el cáncer de mama.

Posteriormente realizamos de nuevo ensayos de MTT para probar un grupo 10 combinaciones fármaco-fármaco y 13 combinaciones nueva diana-fármaco. Se realizó un análisis completo y sistemático de los resultados obtenidos y se evaluó, de cada combinación, los valores DCI50 obtenidos en las seis líneas celulares utilizadas. Los experimentos indican que, considerando sólo DCI50 el 44% de las combinaciones de fármacos interactúan de una manera sinérgica entre sí, el 39% son aditivos y tan solo un 17% son antagónicos. También se computaron los valores de DCI desde el 20% al 80% de respuesta, con el objetivo de mostrar una visión completa del comportamiento de la interacción de los compuestos ensayados en combinación. Todos los detalles de los resultados obtenidos se muestran en el Apéndice 10 de la tesis.

El último análisis de los resultados de MTT fue evaluar el efecto aditivo/sinérgico de las nuevas dianas en combinación, independientemente de los efectos inhibidores individualmente. Interesantemente, siete de los ocho candidatos diana mostraron sinergismo en al menos una combinación. Estos resultados validan fuertemente la estrategia seguida para predecir dichas dianas.

En resumen, los resultados de este capítulo de la tesis demuestran que la estrategia basada en biología de redes es robusta y se puede aplicar a otros tipos de cáncer y enfermedades.

4. Conclusiones

En la primera sección de este trabajo, se muestra cómo una combinación de experimentos de descubrimiento de interacciones y su posterior análisis computacional basado en biología de redes proporciona nuevas evidencias de los posibles roles de los genes causantes y de susceptibilidad relacionados con los cánceres de mama y colorrectal, sugiriendo nuevas hipótesis en cuanto a sus funciones moleculares.

A continuación, hemos propuesto y validado experimentalmente una estrategia de biología de redes para identificar nuevas dianas terapéuticas eficaces, ya sea individualmente o en combinación con medicamentos ya aprobados contra el cáncer de mama. Asimismo, también ha sido desarrollado y evaluado otro enfoque computacional para inferir combinaciones de fármacos teniendo en cuenta la redundancia funcional y la interferencia en la comunicación entre distintas vías de señalización.

En conclusión, las estrategias presentadas en esta tesis ofrecen una perspectiva global para explorar los mecanismos moleculares que subyacen en enfermedades complejas, más allá del estudio de genes individuales. Por lo tanto, dichas estrategias pueden ser aplicadas fácilmente a otros tipos de cáncer y otras enfermedades complejas, así como para el estudio de cuestiones biológicas más específicas relacionadas con el cáncer de mama.

CHAPTER

VIII

APPENDICES

Appendix 1. List of BC causative genes (drivers) and their features

Genes that have been also examined by Y2H library screens are highlighted in orange. Genes depicted in grey were incorrectly cloned.

Symbol	Gene name	UniProt	Molecular weight	Additional information
AKT1	v-akt murine thymoma viral oncogene homolog 1	P31749	55686 Da 480 AA	AKT regulates cell proliferation in breast cancer cells
APC	adenomatous polyposis coli	P25054	311,646Da 2843 AA	genetic variation increases breast cancer
AR	androgen receptor	P10275	98989 Da 919 AA	mutation in the gene in breast cancer patients
ATM	ataxia telangiectasia mutated	Q13315	350644 Da 3056 AA	mutation in the gene associated with breast cancer patients
AURKA	aurora kinase A	O14965	45809 Da 403 AA	overexpressed in human breast cancer cell lines
BAG4	BCL2-associated athanogene 4	O95429	49594 Da 457 AA	breast cancer oncogene
BARD1	BRCA1 associated RING domain 1	Q99728	86648 Da 777 AA	mutation in the gene, evidence of involvement of susceptibility to breast cancer
BCAR3	breast cancer anti-estrogen resistance 3	O75815	92566 Da 825 AA	gene that shows antiestrogen resistance in breast cancer cells
BRCA1	breast cancer 1, early onset	P38398	207721 Da 1863 AA	mutation in the gene, predisposition to breast cancer
BRCA2	breast cancer 2, early onset	P51587	384225 Da 3418 AA	mutation in the gene, predisposition to breast cancer
BRIP1	BRCA1 interacting protein C-terminal helicase 1	Q9BX63	140878 Da 1249 AA	mutation involve in early-onset breast cancer
BRMS1	breast cancer metastasis suppressor 1	Q9HCU9	28461 Da 246 AA	breast carcinoma, metastasis suppressor
CASP8	caspase 8, apoptosis-related	Q14790	55391 Da	associated in breast cancer

Appendix 1. List of BC causative genes (drivers) and their features

	cysteine peptidase		479 AA	
CCND1	cyclin D1	P24385	33729 Da 295 AA	overexpression
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	P12830	97456 Da 882 AA	mutations of the gene in breast cancer
CDKN2A	cyclin-dependent kinase inhibitor 2A	P42771	16533 Da 156 AA	
CDKN2C	cyclin-dependent kinase inhibitor 2C	P42773	18,127Da 168 AA	mutation in the gene
CHEK2	CHK2 checkpoint homolog (S. pombe)	O96017	60915 Da 543 AA	mutation found in breast cancer
E2F1	E2F transcription factor 1	Q01094	46920 Da 437 AA	tumor suppressor gene
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2	P04626	137910 Da 1255 AA	overexpressed
ESR1	estrogen receptor 1	P03372	66216 Da 595 AA	amplification and overexpression in breast cancer cells
ESR2	estrogen receptor 2 (ER beta)	Q92731	59216 Da 530 AA	expressed (T)
FBXW7	F-box and WD repeat domain containing 7	Q969H0	79663 Da 707 AA	
FGFR2	fibroblast growth factor receptor 2	P21802	92025 Da 821 AA	associated with familial breast cancer
FGFR4	fibroblast growth factor receptor 4	P22455	87,954Da 802 AA	mutation in the gene implicated in cancer progression and tumor cell motility
HMMR	hyaluronan-mediated motility receptor	O75330	84100 Da 724 AA	association in breast tumorigenesis between BRCA1 and HMMR with AURKA
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	P01112	21,298Da 189 AA	polymorphism
IGF1R	insulin-like growth factor 1 receptor	P08069	154793 Da 1367 AA	substantial changes of the gene in breast tumorigenesis

KRAS	GTPase KRas	P01116	21656 Da 189 AA	mutations found in the gene
LSP1	lymphocyte-specific protein 1	P33241	37,192Da 339 AA	associated with familial breast cancer
MAP3K1	mitogen-activated protein kinase kinase kinase 1	Q13233	164470 Da 1512 AA	associated with familial breast cancer
MAPK14	mitogen-activated protein kinase 14	Q16539	41293 Da 360 AA	mediates cell cycle progression in breast cancer
NBN	nibrin	O60934	84959 Da 754 AA	mutation
NCOA3	nuclear receptor coactivator 3	Q9Y6Q9	155293 Da 1424 AA	overexpressed (T)
NOTCH2	Neurogenic locus notch homolog protein 2	Q04721	265405 Da 2471 AA	tumor suppressive in breast cancer
PALB2	partner and localizer of BRCA2	Q86YC2	131523 Da 1186 AA	mutation
PARP1	poly (ADP-ribose) polymerase 1	P09874	113,084Da, 1014 AA	mutations
PAX2	paired box 2	Q02962	44734 Da 416 AA	expressed (T)
PGR	progesterone receptor	P06401	98981 Da 933 AA	Cell movement and invasion
PHB	prohibitin	P35232	29804 Da 272 AA	mutations
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide	P42336	124284 Da 1068 AA	mutation
PPM1D	protein phosphatase, Mg2+/Mn2+ dependent, 1D	O15297	66675 Da 605 AA	Amplify (CL)
PTEN	phosphatase and tensin homolog	P60484	47166 Da 403 AA	familial breast cancer
PTPN1	protein tyrosine phosphatase, non-receptor type 1	P18031	49967 Da 435 AA	overexpression of PTP1B in mammary gland led to spontaneous breast cancer development

Appendix 1. List of BC causative genes (drivers) and their features

PTPRJ	protein tyrosine phosphatase, receptor type, J	Q12913	145941 Da 1337 AA	frequent deletions, loss of heterozygosity, and missense mutations
RAD50	RAD50 homolog	Q92878	153892 Da 1312 AA	mutations
RAD51	RAD51 homolog	Q06609	36966 Da 339 AA	mutations
RAD54L	RAD54-like	Q92698	84352 Da 747 AA	LOH
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	P04049	73052 Da 648 AA	overexpressed
RB1CC1	RB1-inducible coiled-coil 1	Q8TDY2	183091 Da 1594 AA	(20%) primary breast cancers examined contained mutations in RB1CC1
SMAD4	SMAD family member 4	Q13485	60439 Da 552 AA	homozygous deletion
STK11	serine/threonine kinase 11	Q15831	48636 Da 433 AA	loss of expresion
TGFB1	transforming growth factor, beta 1	P01137	44341 Da 390 AA	can promote the formation of lung metastases
TGFBR1	transforming growth factor, beta receptor 1	P36897	55960 Da 530 AA	enhances the migration and invasion
TGFBR3	transforming growth factor, beta receptor III	Q03167	93428 Da 850 AA	loss of expresion
TP53	tumor protein p53	P04637	43712 Da 393 AA	polymorphic variants
TSG101	tumor susceptibility gene 101	Q99816	43944 Da 390 AA	mutated
WT1	Wilms tumor 1	P19544	49188 Da 449 AA	strongly expressed in primary carcinomas
XRCC3	X-ray repair complementing defective repair 3	O43542	37850 Da 346 AA	mutations

Appendix 2. List of BC candidate genes and their features

Network level of each gene is indicated (only 'third_level' and 'not_present' genes were selected).
Grey = not correctly cloned into prey plasmid. T= tumor tissue, CL= cell lines

Symbol	Gene name	UniProt	Molecular weight	Network level	Additional information
AGR3	anterior gradient homolog 3	Q8TD06	19171 Da 166 AA	not present	expressed in breast cancer tissues
ANGPTL4	angiopoietin-like 4	Q9BY76	45,214Da 406 AA	not present	Genes that mediate BC metastasis to lung
ATAD2	ATPase family, AAA domain containing 2	Q6PL18	158554 Da 1390 AA	not present	gene upregulated in BC cells overexpressing ACTR or treated with estrogen
BAP1	BRCA1 associated protein-1	Q92560	80362 Da 729 AA	not present	BAP1 enhances BRCA1-mediated inhibition of breast cancer cell growth
BCAS3	breast carcinoma amplified sequence 3	Q9H6U6	101237 Da 928 AA	not present	amplification and overexpression in breast cancer cells
BEX1	brain expressed, X-linked 1	Q9HBH7	14,860Da 125 AA	not present	overexpressed in human breast cancer cell lines
BEX2	brain expressed X-linked 2	Q9BXY8	15321 Da 128 AA	third level	overexpressed in human breast cancer cell lines
BLID	BH3-like motif-containing cell death inducer	Q8IZY5	12045 Da 108 AA	not present	Amplify from a breast cancer cell line
C3orf35 (APRG1)	chromosome 3 open reading frame 35	Q8IVJ8	18525 Da 170 AA	not present	reduced (T)
CASC3	cancer susceptibility candidate 3	O15234	76278Da 703 AA	third level	overexpressed in breast carcinoma cell line
CASZ1	castor zinc finger 1	Q86V15	190148 Da 1759 AA	not present	present in the perinuclear regions of cancer cells

Appendix 2. List of BC candidate genes and their features

CCL5	chemokine (C-C motif) ligand 5	P13501	9,990 91 AA	third level	secreted in breast cancer (T)
CDH13	cadherin 13, H-cadherin (heart)	P55290	78287 Da 713 AA	not present	downregulated (T)
CST6	cystatin E/M	Q15828	16511 Da 149 AA	not present	downregulated (T)
CTCFL	CCCTC-binding factor (zinc finger protein)-like	Q8NI51	75717 Da 663 AA	not present	expressed (T)
CXCL1	chemokine (C-X-C motif) ligand 1	P09341	11,301Da 107 AA	third level	metastasis to the lungs
CYP17A1	cytochrome P450, family 17, subfamily A, polypeptide 1	P05093	57371 Da 508 AA	third level	mutation in breast cancer
DIRAS3	DIRAS family, GTP-binding RAS-like 3	O95661	25,861Da 229 AA	not present	downregulated (T)
DKK3	dickkopf homolog 3	Q9UBP4	38291 Da 350 AA	third level	downregulated (T)
EPSTI1	epithelial stromal interaction 1 (breast)	Q96J88	36793 Da 318 AA	not present	upregulated (T)
ERRFI1	ERBB receptor feedback inhibitor 1	Q9UJM3	50560 Da 462 AA	not present	downregulated (T)
FAM84B	family with sequence similarity 84, member B	Q96KN1	34474 Da 310 AA	not present	upregulated (T)
GLCE	glucuronic acid epimerase	O94923	70,115Da 617 AA	not present	lower GLCE mRNA and protein in most BC tumor tissue and surrounding nontumor tissue
GREB1	growth regulation by estrogen in breast cancer 1	Q4ZG55	216467 Da 1949 AA	not present	overexpressed in ER+ (T)
HOXC6	homeobox C6	P09630	26915 Da 235 AA	not present	downregulated (CL)
IL13RA2	interleukin 13 receptor, alpha 2	Q14627	44,176Da 380 AA	third level	Gene that mediate breast cancer metastasis to lung
IL24	interleukin 24	Q13007	23825 Da 206 AA	not present	induces apoptosis in human breast cancer cells

ITIH5	inter-alpha (globulin) inhibitor H5	Q86UX2	104576 Da 942 AA	not present	metastasize lung
KIAA0100	KIAA0100	Q14667	253700 Da 2235 AA	not present	overexpressed (T)
KLK5	kallikrein-related peptidase 5	Q9Y337	32020 Da 293 AA	third level	upregulated (T)
KLK6	kallikrein-related peptidase 6	Q92876	26856 Da 244 AA	third level	overexpressed (but not in the corresponding metastatic cell lines from the same patient)
KLK7	kallikrein-related peptidase 7	P49862	27525 Da 253 AA	not present	upregulation
KLK9	kallikrein-related peptidase 9	Q9UKQ9	27513 Da 250 AA	not present	KLK9 gene is regulated by steroid hormones in a human breast cancer cell line
LYPD3	LY6/PLAUR domain containing 3	O95274	35971 Da 346 AA	not present	overexpressed (T)
MRC2	mannose receptor, C type 2	Q9UBG0	166655 Da 1479 AA	not present	high levels in breast cancer (T)
MTA3	metastasis associated 1 family, member 3	Q9BTC8	67,504Da 594 AA	not present	regulates an invasive growth pathway in breast cancer
NAT2	N-acetyltransferase 2	P11245	33542 Da 290 AA	not present	associated
OSGIN1	oxidative stress induced growth inhibitor 1	Q9UJX0	60,820Da 560 AA	not present	inhibits growth and migration of breast cancer cells
PDLIM2	PDZ and LIM domain 2	Q96JY6	37459 Da 352 AA	not present	overexpressed (T)
PRDM14	PR domain containing 14	Q9GZV8	64062 Da 571 AA	not present	overexpression (T)
PSMC3IP	PSMC3 interacting protein	Q9P2W1	24906 Da 217 AA	not present	high expression (T)
RASL10B	RAS-like, family 10, member B	Q96S79	23,229Da 203 AA	not present	downregulated (CL)
RHOBTB2	Rho-related BTB domain containing 2	Q9BYZ6	82626 Da 727 AA	not present	homozygously deleted

Appendix 2. List of BC candidate genes and their features

RNF20	ring finger protein 20	Q5VTR2	113662 Da 975 AA	not present	hypermethylated (downregulated)
S100A14	S100 calcium binding protein A14	Q9HCY8	11662 Da 104 AA	not present	overexpressed (T)
SCGB2A2	secretoglobin, family 2A, member 2	Q13296	10499 Da 93 AA	not present	overexpressed (T)
SCGB3A1	secretoglobin, family 3A, member 1	Q96QR1	10100 Da 104 AA	not present	loss of expression (CL)
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	P36952	42138 Da 375 AA	not present	tumos suppresin
SNAI1	snail homolog 1	O95863	29083 Da 264 AA	third level	downregulated (T)
ST14	suppression of tumorigenicity 14	Q9Y5Y6	94770 Da 855 AA	not present	LOH
TFF1	trefoil factor 1	P04155	9150 Da 84 AA	third level	detected in approximately 50% of human breast tumors
THRSP	thyroid hormone responsive	Q92748	16561 Da 146 AA	not present	overexpressed (T)
TRIM25	tripartite motif-containing 25	Q14258	70989 Da 630 AA	third level	expressed in ER+ (T)
TSP50	protease, serine, 50	Q9UI38	43088 Da 385 AA	not present	up-regulated (T)
VPS45	vacuolar protein sorting 45 homolog	Q9NRW7	65077 Da 570 AA	third level	not expressed
WHSC1L1	Wolf-Hirschhorn syndrome candidate 1-like 1	Q9BZ95	161613 Da 1437 AA	not present	amplified
WIF1	WNT inhibitory factor 1	Q9Y5W5	41528 Da 379 AA	not present	breast tumor cell lines display absent or low levels of WIF1 expression
ZNF217	zinc finger protein 217	O75362	115272 Da 1048 AA	not present	overexpressed (CL)

Appendix 3. List of interactions associated to BC detected by yeast two-hybrid matrix screens

Interacting pairs are reported as gene symbols taken from HGNC (http://www.genenames.org). The confidence level (high or low) is reported.

HIGH CONFIDENCE

	Bait	Prey
1	AKT1	AGR3
2	AKT1	CASC3
3	AKT1	EPSTI1
4	AKT1	IL13RA2
5	AKT1	IL24
6	AKT1	MRC2
7	AKT1	MTA3
8	AKT1	OSGIN1
9	AKT1	RASL10B
10	AKT1	S100A14
11	AKT1	SERPINB5
12	AKT1	ST14
13	AKT1	TFF1
14	AKT1	THRSP
15	AKT1	WHSC1L1
16	APC	AGR3
17	APC	CASC3
18	APC	CCL5
19	APC	CYP17A1
20	APC	DIRAS3
21	APC	DKK3
22	APC	HOXC6
23	APC	IL24
24	APC	NAT2
25	APC	PDLIM2
26	APC	ST14
27	APC	TFF1
28	APC	TRIM25
29	ATM	BCAS3
30	ATM	ERRFI1
31	ATM	IL24
32	ATM	MTA3
33	ATM	NAT2
34	ATM	OSGIN1
35	ATM	TFF1
36	ATM	WHSC1L1
37	AURKA	ANGPTL4
38	AURKA	BEX2
39	AURKA	BLID
40	AURKA	CDH13
41	AURKA	DKK3
42	AURKA	EPSTI1
43	AURKA	ERRFI1
44	AURKA	KLK5
45	AURKA	LYPD3
46	AURKA	MTA3
47	AURKA	NAT2
48	AURKA	PDLIM2
49	AURKA	PSMC3IP
50	AURKA	S100A14
51	AURKA	SCGB3A1
52	AURKA	THRSP
53	AURKA	WIF1
54	BAG4	BEX1
55	BAG4	IL24
56	BAG4	LYPD3
57	BAG4	BCAS3
58	BAG4	NAT2
59	BAG4	ATAD2
60	BAG4	BAP1
61	BAG4	KLK5
62	BAG4	RNF20
63	BAG4	SCGB3A1
64	BAG4	THRSP
65	BAG4	CASZ1
66	BAG4	CCL5
67	BAG4	DIRAS3
68	BAG4	EPSTI1
69	BAG4	FAM84B
70	BAG4	GLCE
71	BAG4	IL13RA2
72	BAG4	KLK9
73	BAG4	PDLIM2
74	BAG4	ERRFI1
75	BAG4	KLK7
76	BAG4	MTA3
77	BAG4	PRDM14
78	BAG4	RASL10B
79	BAG4	ITIH5
80	BAG4	MRC2
81	BAG4	RHOBTB2
82	BCAR3	AGR3
83	BCAR3	ATAD2

Appendix 3. List of interactions associated to BC detected by yeast two-hybrid matrix screens

84	BCAR3	BAP1
85	BCAR3	BEX1
86	BCAR3	BEX2
87	BCAR3	CASZ1
88	BCAR3	CCL5
89	BCAR3	CST6
90	BCAR3	DKK3
91	BCAR3	EPSTI1
92	BCAR3	ERRFI1
93	BCAR3	GLCE
94	BCAR3	GREB1
95	BCAR3	HOXC6
96	BCAR3	MTA3
97	BCAR3	OSGIN1
98	BCAR3	PDLIM2
99	BCAR3	PRDM14
100	BCAR3	PSMC3IP
101	BCAR3	SERPINB5
102	BCAR3	THRSP
103	BCAR3	VPS45
104	BCAR3	WHSC1L1
105	BRMS1	ANGPTL4
106	BRMS1	BEX2
107	BRMS1	DIRAS3
108	BRMS1	IL13RA2
109	BRMS1	KLK9
110	BRMS1	LYPD3
111	BRMS1	MRC2
112	BRMS1	PSMC3IP
113	BRMS1	SNAI1
114	BRMS1	THRSP
115	CASP8	AGR3
116	CASP8	ANGPTL4
117	CASP8	ATAD2
118	CASP8	BAP1
119	CASP8	BEX1
120	CASP8	BLID
121	CASP8	CASC3
122	CASP8	CDH13
123	CASP8	EPSTI1
124	CASP8	ERRFI1
125	CASP8	GLCE
126	CASP8	GREB1
127	CASP8	IL13RA2
128	CASP8	IL24
129	CASP8	KLK5
130	CASP8	KLK9
131	CASP8	LYPD3
132	CASP8	PRDM14
133	CASP8	PSMC3IP
134	CASP8	S100A14
135	CASP8	SCGB2A2
136	CASP8	SERPINB5
137	CASP8	ST14
138	CASP8	TSP50
139	CASP8	VPS45
140	CASP8	WHSC1L1
141	CCND1	BCAS3
142	CCND1	CDH13
143	CCND1	KLK7
144	CCND1	KLK9
145	CDH1	BEX1
146	CDH1	BLID
147	CDH1	CCL5
148	CDH1	CYP17A1
149	CDH1	DKK3
150	CDH1	EPSTI1
151	CDH1	MTA3
152	CDH1	NAT2
153	CDH1	PSMC3IP
154	CDH1	THRSP
155	CDKN2A	ANGPTL4
156	CDKN2A	CASC3
157	CDKN2A	NAT2
158	CDKN2A	WHSC1L1
159	CDKN2C	ANGPTL4
160	CDKN2C	GREB1
161	CDKN2C	MTA3
162	CDKN2C	NAT2
163	CDKN2C	SCGB2A2
164	CDKN2C	ST14
165	CHEK2	BEX1
166	CHEK2	CCL5
167	CHEK2	CDH13
168	CHEK2	CYP17A1
169	CHEK2	DIRAS3
170	CHEK2	FAM84B
171	CHEK2	HOXC6
172	CHEK2	IL24
173	CHEK2	ITIH5
174	CHEK2	KLK7
175	CHEK2	KLK9
176	CHEK2	LYPD3
177	CHEK2	PRDM14
178	CHEK2	PSMC3IP
179	CHEK2	RHOBTB2
180	CHEK2	RNF20
181	CHEK2	ST14
182	CHEK2	WHSC1L1
183	ERBB2	CYP17A1
184	ERBB2	ERRFI1
185	ERBB2	GLCE
186	ERBB2	IL13RA2
187	ERBB2	KLK5
188	ERBB2	LYPD3
189	ERBB2	NAT2
190	ERBB2	PRDM14
191	ERBB2	PSMC3IP
192	ERBB2	RHOBTB2
193	ERBB2	THRSP
194	ESR1	ANGPTL4
195	ESR1	ATAD2
196	ESR1	BAP1
197	ESR1	CXCL1
198	ESR1	EPSTI1
199	ESR1	GLCE

200	ESR1	KLK5
201	ESR1	KLK9
202	ESR1	PSMC3IP
203	ESR1	SERPINB5
204	ESR1	WHSC1L1
205	ESR2	AGR3
206	ESR2	CASC3
207	ESR2	CASZ1
208	ESR2	DKK3
209	ESR2	FAM84B
210	ESR2	IL13RA2
211	ESR2	IL24
212	ESR2	ITIH5
213	ESR2	KLK9
214	ESR2	MRC2
215	ESR2	TFF1
216	ESR2	TSP50
217	FBXW7	ANGPTL4
218	FBXW7	BCAS3
219	FBXW7	BEX1
220	FBXW7	IL24
221	FBXW7	SCGB3A1
222	FBXW7	TRIM25
223	FGFR2	BEX1
224	FGFR2	BEX2
225	FGFR2	ERRFI1
226	FGFR2	GLCE
227	FGFR2	HOXC6
228	FGFR2	MTA3

229	FGFR2	PDLIM2
230	FGFR2	RASL10B
231	FGFR2	RHOBTB2
232	FGFR2	S100A14
233	FGFR2	TFF1
234	FGFR4	ANGPTL4
235	FGFR4	BLID
236	FGFR4	CASC3
237	FGFR4	DIRAS3
238	FGFR4	EPSTI1
239	FGFR4	ERRFI1
240	FGFR4	FAM84B
241	FGFR4	GLCE
242	FGFR4	MTA3
243	FGFR4	OSGIN1
244	FGFR4	PDLIM2
245	FGFR4	PSMC3IP
246	FGFR4	RASL10B
247	FGFR4	RNF20
248	FGFR4	SCGB3A1
249	FGFR4	THRSP
250	FGFR4	WHSC1L1
251	FGFR4	WIF1
252	HMMR	CDH13
253	HMMR	GREB1
254	HMMR	IL24
255	HMMR	ITIH5
256	HMMR	NAT2
257	HMMR	THRSP

258	HRAS	BLID
259	HRAS	CXCL1
260	HRAS	GREB1
261	HRAS	IL24
262	HRAS	TSP50
263	IGF1R	BAP1
264	IGF1R	CYP17A1
265	IGF1R	HOXC6
266	IGF1R	IL13RA2
267	IGF1R	KLK5
268	IGF1R	LYPD3
269	IGF1R	MTA3
270	IGF1R	NAT2
271	IGF1R	THRSP
272	IGF1R	TRIM25
273	IGF1R	VPS45
274	KRAS	BCAS3
275	KRAS	CCL5
276	KRAS	DIRAS3
277	KRAS	IL24
278	KRAS	OSGIN1
279	KRAS	SERPINB5
280	KRAS	THRSP
281	LSP1	GREB1
282	LSP1	IL24
283	LSP1	KLK6
284	LSP1	LYPD3
285	LSP1	SCGB2A2
286	LSP1	SNAI1

287	LSP1	THRSP
288	NBN	BAP1
289	NBN	CASC3
290	NBN	NAT2
291	NBN	SNAI1
292	NOTCH2	CST6
293	NOTCH2	EPSTI1
294	NOTCH2	FAM84B
295	NOTCH2	IL13RA2
296	NOTCH2	IL24
297	NOTCH2	ITIH5
298	NOTCH2	KLK5
299	NOTCH2	MTA3
300	NOTCH2	PSMC3IP
301	NOTCH2	ST14
302	PALB2	BCAS3
303	PALB2	CASZ1
304	PALB2	CCL5
305	PALB2	CST6
306	PALB2	DIRAS3
307	PALB2	ERRFI1
308	PALB2	FAM84B
309	PALB2	GREB1
310	PALB2	IL13RA2
311	PALB2	IL24
312	PALB2	KLK6
313	PALB2	LYPD3
314	PALB2	MRC2
315	PALB2	NAT2

Appendix 3. List of interactions associated to BC detected by yeast two-hybrid matrix screens

316	PALB2	PRDM14
317	PALB2	RHOBTB2
318	PALB2	SERPINB5
319	PALB2	SNAI1
320	PALB2	THRSP
321	PARP1	BLID
322	PARP1	IL24
323	PARP1	MTA3
324	PARP1	RASL10B
325	PARP1	THRSP
326	PAX2	BEX1
327	PAX2	BLID
328	PAX2	GLCE
329	PAX2	IL13RA2
330	PAX2	KLK6
331	PAX2	KLK9
332	PAX2	MRC2
333	PAX2	PRDM14
334	PAX2	RASL10B
335	PAX2	SNAI1
336	PAX2	TRIM25
337	PAX2	VPS45
338	PAX2	WHSC1L1
339	PHB	BCAS3
340	PHB	CASC3
341	PHB	CCL5
342	PHB	ST14
343	PIK3CA	BEX1
344	PIK3CA	BEX2

345	PIK3CA	IL13RA2
346	PIK3CA	IL24
347	PIK3CA	PSMC3IP
348	PIK3CA	THRSP
349	PPM1D	BEX2
350	PPM1D	CASZ1
351	PPM1D	DIRAS3
352	PPM1D	DKK3
353	PPM1D	EPSTI1
354	PPM1D	ERRFI1
355	PPM1D	GLCE
356	PPM1D	ITIH5
357	PPM1D	LYPD3
358	PPM1D	MRC2
359	PPM1D	NAT2
360	PPM1D	OSGIN1
361	PPM1D	RHOBTB2
362	PPM1D	WHSC1L1
363	PTEN	BEX1
364	PTEN	CXCL1
365	PTEN	IL24
366	PTEN	OSGIN1
367	PTPN1	AGR3
368	PTPN1	BCAS3
369	PTPN1	CASC3
370	PTPN1	CASZ1
371	PTPN1	CCL5
372	PTPN1	CDH13
373	PTPN1	DIRAS3

374	PTPN1	FAM84B
375	PTPN1	KLK7
376	PTPN1	MRC2
377	PTPN1	RHOBTB2
378	PTPN1	SNAI1
379	PTPN1	THRSP
380	PTPN1	VPS45
381	PTPRJ	BAP1
382	PTPRJ	CASC3
383	PTPRJ	CYP17A1
384	PTPRJ	EPSTI1
385	PTPRJ	HOXC6
386	PTPRJ	KLK7
387	PTPRJ	KLK9
388	PTPRJ	LYPD3
389	PTPRJ	PRDM14
390	PTPRJ	TRIM25
391	RAD51	CDH13
392	RAD51	CST6
393	RAD51	FAM84B
394	RAD51	IL24
395	RAD51	ITIH5
396	RAD51	NAT2
397	RAD51	ST14
398	RAD51	TFF1
399	RAD51	WHSC1L1
400	RAF1	CASZ1
401	RAF1	MRC2
402	RAF1	PDLIM2

403	RB1CC1	BAP1
404	RB1CC1	BLID
405	RB1CC1	EPSTI1
406	RB1CC1	ERRFI1
407	RB1CC1	GLCE
408	RB1CC1	GREB1
409	RB1CC1	IL13RA2
410	RB1CC1	IL24
411	RB1CC1	ITIH5
412	RB1CC1	LYPD3
413	RB1CC1	MRC2
414	RB1CC1	PSMC3IP
415	RB1CC1	RNF20
416	RB1CC1	THRSP
417	RB1CC1	VPS45
418	RB1CC1	WHSC1L1
419	RB1CC1	WIF1
420	SMAD4	BCAS3
421	SMAD4	BEX1
422	SMAD4	DKK3
423	SMAD4	FAM84B
424	SMAD4	GREB1
425	SMAD4	IL24
426	SMAD4	LYPD3
427	SMAD4	S100A14
428	SMAD4	SERPINB5
429	SMAD4	THRSP
430	STK11	BCAS3
431	STK11	CYP17A1

432	STK11	KLK7
433	STK11	LYPD3
434	STK11	NAT2
435	STK11	RHOBTB2
436	STK11	S100A14
437	STK11	SERPINB5
438	STK11	ST14
439	STK11	THRSP
440	STK11	TRIM25
441	STK11	TSP50
442	TGFB1	AGR3
443	TGFB1	ANGPTL4
444	TGFB1	BCAS3
445	TGFB1	BLID
446	TGFB1	CCL5
447	TGFB1	CST6
448	TGFB1	DIRAS3
449	TGFB1	GREB1
450	TGFB1	HOXC6
451	TGFB1	KLK5
452	TGFB1	KLK7
453	TGFB1	KLK9
454	TGFB1	MTA3
455	TGFB1	PDLIM2
456	TGFB1	PSMC3IP
457	TGFB1	SCGB3A1
458	TGFB1	SNAI1
459	TGFB1	ST14
460	TGFB1	THRSP

461	TGFB1	TSP50
462	TGFB1	WHSC1L1
463	TSG101	BCAS3
464	TSG101	KLK5
465	TSG101	KLK6
466	TSG101	MTA3
467	TSG101	RASL10B
468	TSG101	SCGB3A1
469	TSG101	TFF1
470	TSG101	TSP50
471	WT1	CASZ1
472	WT1	CCL5
473	WT1	DIRAS3
474	WT1	KLK7
475	WT1	NAT2
476	WT1	THRSP
477	XRCC3	BAP1
478	XRCC3	BCAS3
479	XRCC3	CXCL1
480	XRCC3	FAM84B
481	XRCC3	GREB1
482	XRCC3	ITIH5
483	XRCC3	MRC2
484	XRCC3	OSGIN1
485	XRCC3	PSMC3IP
486	XRCC3	RNF20
487	XRCC3	SCGB3A1
488	XRCC3	SERPINB5
489	XRCC3	SNAI1

490	XRCC3	TFF1
491	XRCC3	TRIM25

LOW CONFIDENCE

	Bait	Prey
492	AKT1	BAP1
493	AKT1	CCL5
494	AKT1	TRIM25
495	APC	FAM84B
496	APC	GLCE
497	APC	PRDM14
498	APC	TSP50
499	ATM	BEX2
500	ATM	CASZ1
501	ATM	CXCL1
502	ATM	LYPD3
503	ATM	PRDM14
504	ATM	RNF20
505	ATM	THRSP
506	ATM	TRIM25
507	AURKA	ATAD2
508	AURKA	GLCE
509	AURKA	KLK9
510	AURKA	TSP50
511	BAG4	BEX2
512	BAG4	CDH13
513	BAG4	CST6

514	BAG4	CXCL1
515	BAG4	DKK3
516	BAG4	KLK6
517	BAG4	PSMC3IP
518	BAG4	S100A14
519	BAG4	ST14
520	BAG4	TSP50
521	BAG4	WIF1
522	BCAR3	CXCL1
523	BCAR3	DIRAS3
524	BCAR3	KLK6
525	BCAR3	MRC2
526	BCAR3	NAT2
527	BCAR3	SNAI1
528	BRMS1	ATAD2
529	BRMS1	ITIH5
530	BRMS1	PDLIM2
531	BRMS1	ST14
532	BRMS1	TSP50
533	BRMS1	VPS45
534	CASP8	CASZ1
535	CASP8	MRC2
536	CASP8	RHOBTB2
537	CASP8	TFF1
538	CCND1	CASZ1
539	CCND1	TFF1
540	CDH1	FAM84B
541	CDH1	KLK6
542	CDKN2A	BCAS3

Appendix 3. List of interactions associated to BC detected by yeast two-hybrid matrix screens

543	CDKN2A	KLK9
544	CDKN2A	MTA3
545	CDKN2A	VPS45
546	CDKN2C	CASZ1
547	CDKN2C	EPSTI1
548	CDKN2C	MRC2
549	CHEK2	BAP1
550	CHEK2	BEX2
551	CHEK2	CST6
552	CHEK2	EPSTI1
553	CHEK2	NAT2
554	CHEK2	TSP50
555	ERBB2	CASZ1
556	ERBB2	MRC2
557	ERBB2	SERPINB5
558	ESR1	CYP17A1
559	ESR1	NAT2
560	ESR1	THRSP
561	ESR1	TSP50
562	ESR1	VPS45
563	ESR1	WIF1
564	ESR2	CXCL1
565	ESR2	SNAI1
566	ESR2	THRSP
567	ESR2	WHSC1L1
568	FBXW7	CASZ1
569	FBXW7	DIRAS3
570	FBXW7	NAT2
571	FGFR2	ANGPTL4
572	FGFR2	BAP1
573	FGFR2	CCL5
574	FGFR2	KLK5
575	FGFR2	KLK6
576	FGFR2	KLK9
577	FGFR2	SNAI1
578	FGFR4	AGR3
579	FGFR4	BEX2
580	FGFR4	CXCL1
581	FGFR4	KLK5
582	FGFR4	PRDM14
583	FGFR4	S100A14
584	FGFR4	ST14
585	FGFR4	TRIM25
586	HMMR	ANGPTL4
587	HMMR	BCAS3
588	HMMR	KLK5
589	HMMR	MRC2
590	HMMR	RASL10B
591	HMMR	SNAI1
592	HRAS	DKK3
593	HRAS	NAT2
594	HRAS	VPS45
595	IGF1R	BCAS3
596	IGF1R	BLID
597	IGF1R	CASC3
598	IGF1R	CXCL1
599	IGF1R	DIRAS3
600	IGF1R	ERRFI1
601	IGF1R	GLCE
602	IGF1R	MRC2
603	IGF1R	PDLIM2
604	IGF1R	RASL10B
605	IGF1R	S100A14
606	IGF1R	SCGB3A1
607	KRAS	CASZ1
608	KRAS	MRC2
609	KRAS	SCGB2A2
610	KRAS	TRIM25
611	KRAS	WHSC1L1
612	LSP1	BLID
613	LSP1	CCL5
614	LSP1	VPS45
615	NBN	BCAS3
616	NBN	BEX2
617	NBN	HOXC6
618	NBN	KLK6
619	NBN	PRDM14
620	NBN	RASL10B
621	NBN	TRIM25
622	NOTCH2	BCAS3
623	NOTCH2	CASZ1
624	NOTCH2	CYP17A1
625	NOTCH2	DKK3
626	NOTCH2	GREB1
627	NOTCH2	LYPD3
628	NOTCH2	PDLIM2
629	NOTCH2	RHOBTB2
630	NOTCH2	SNAI1
631	NOTCH2	THRSP
632	PALB2	BAP1
633	PALB2	BEX1
634	PALB2	ITIH5
635	PALB2	TRIM25
636	PALB2	WHSC1L1
637	PARP1	AGR3
638	PARP1	CDH13
639	PARP1	DKK3
640	PARP1	FAM84B
641	PARP1	ITIH5
642	PARP1	KLK7
643	PARP1	NAT2
644	PARP1	PRDM14
645	PARP1	S100A14
646	PARP1	ST14
647	PARP1	TFF1
648	PARP1	TSP50
649	PAX2	KLK7
650	PAX2	NAT2
651	PAX2	THRSP
652	PHB	DKK3
653	PHB	RHOBTB2
654	PHB	SCGB3A1
655	PIK3CA	BCAS3
656	PIK3CA	FAM84B
657	PIK3CA	ITIH5
658	PPM1D	BEX1

659	PPM1D	KLK6
660	PPM1D	PSMC3IP
661	PPM1D	SERPINB5
662	PPM1D	VPS45
663	PTEN	SERPINB5
664	PTPN1	GREB1
665	PTPN1	PSMC3IP
666	PTPN1	TRIM25
667	PTPRJ	BCAS3
668	PTPRJ	IL13RA2
669	PTPRJ	NAT2
670	PTPRJ	SCGB3A1
671	RAD51	BCAS3
672	RAF1	EPSTI1
673	RAF1	SERPINB5
674	RAF1	THRSP
675	RAF1	VPS45
676	RB1CC1	AGR3
677	RB1CC1	ANGPTL4
678	RB1CC1	BEX1
679	RB1CC1	DKK3
680	RB1CC1	NAT2
681	RB1CC1	RASL10B
682	RB1CC1	RHOBTB2
683	RB1CC1	S100A14
684	RB1CC1	TSP50
685	SMAD4	CXCL1
686	SMAD4	DIRAS3
687	SMAD4	KLK7
688	SMAD4	NAT2
689	STK11	ANGPTL4
690	STK11	BEX1
691	STK11	CASZ1
692	STK11	CST6
693	STK11	DKK3
694	STK11	ITIH5
695	STK11	MRC2
696	STK11	MTA3
697	STK11	SCGB2A2
698	STK11	TFF1
699	TGFB1	BEX1
700	TGFB1	CDH13
701	TGFB1	CYP17A1
702	TGFB1	EPSTI1
703	TGFB1	GLCE
704	TGFB1	NAT2
705	TGFB1	RASL10B
706	TGFB1	RHOBTB2
707	TGFB1	S100A14
708	TGFB1	SERPINB5
709	TSG101	ATAD2
710	TSG101	CASZ1
711	TSG101	DIRAS3
712	TSG101	SERPINB5
713	TSG101	THRSP
714	WT1	BLID
715	WT1	FAM84B
716	WT1	KLK9
717	WT1	LYPD3
718	WT1	MRC2
719	WT1	MTA3
720	WT1	SCGB2A2
721	WT1	WHSC1L1
722	XRCC3	AGR3
723	XRCC3	BLID
724	XRCC3	DIRAS3
725	XRCC3	DKK3
726	XRCC3	KLK6
727	XRCC3	PDLIM2
728	XRCC3	ST14

Appendix 4. List of interactions associated to BC detected by yeast two-hybrid library screens

Interacting pairs are reported as gene symbols taken from HGNC (http://www.genenames.org).

	Bait symbol	Prey symbol	Prey gene name	Blast link (Gene ID)	Prey Uniprot
42	CHEK2	SLC25A6	Homo sapiens solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6 (SLC25A6), nuclear gene encoding mitochondrial protein	NM_001636.2	P12236
1	AKT1	ACAT2	ACAT2 acetyl-Coenzyme A acetyltransferase 2	39	O75908
46	ERBB2	ANXA6	ANXA6 annexin A6	309	P08133
87	STK11	ARG2	ARG2 arginase, type II	384	P78540
30	CASP8	ATP1A3	ATP1A3 ATPase, Na+/K+ transporting, alpha 3 polypeptide	478	P13637
59	ERBB2	ATP6AP1	ATP6AP1 ATPase, H+ transporting, lysosomal accessory protein 1	537	ATP6AP1
47	ERBB2	BAI1	BAI1 brain-specific angiogenesis inhibitor 1	575	O94812
88	STK11	CALM2	CALM2 calmodulin 2 (phosphorylase kinase, delta)	805	P62158
62	ESR1	CFL1	CFL1 cofilin 1 (non-muscle)	1072	P23528
89	STK11	CFL2	CFL2 cofilin 2 (muscle)	1073	Q9Y281
48	ERBB2	CLCN6	CLCN6 chloride channel 6	1185	P51797
90	STK11	CRY2	CRY2 cryptochrome 2 (photolyase-like)	1408	Q49AN0
75	RAD51	CSNK2B	CSNK2B casein kinase 2, beta polypeptide	1460	P67870
63	ESR1	CSRP1	CSRP1 cysteine and glycine-rich protein 1	1465	P21291
91	STK11	CTSB	CTSB cathepsin B	1508	P07858

93	STK11	EEF1A2	EEF1A2 eukaryotic translation elongation factor 1 alpha 2	1917	Q05639
39	CHEK2	ENO1	ENO1 enolase 1, (alpha)	2023	P06733
5	AKT1	ENO2	ENO2 enolase 2 (gamma, neuronal)	2026	P09104
6	AKT1	FASN	FASN fatty acid synthase	2194	P49327
23	ATM	FECH	FECH ferrochelatase (protoporphyria)	2235	P22830
7	AKT1	GFAP	GFAP glial fibrillary acidic protein	2670	P14136
40	CHEK2	GNAS	GNAS GNAS complex locus	2778	P63092
8	AKT1	GRIN2A	GRIN2A glutamate receptor, ionotropic, N-methyl D-aspartate 2A	2903	Q12879
50	ERBB2	HDLBP	HDLBP high density lipoprotein binding protein	3069	Q00341
41	CHEK2	HMGN1	HMGN1 high-mobility group nucleosome binding domain 1	3150	P05114
27	ATM	NR4A1	NR4A1 nuclear receptor subfamily 4, group A, member 1	3164	P22736
43	CHEK2	PRMT2	PRMT2 protein arginine methyltransferase 2	3275	P55345
24	ATM	HSPA8	HSPA8 heat shock 70kDa protein 8	3312	P11142
76	RAD51	HSP90AA1	HSP90AA1 heat shock protein 90kDa alpha, class A member 1	3320	P07900
51	ERBB2	IARS	IARS isoleucyl-tRNA synthetase	3376	P41252
44	CHEK2	STMN1	STMN1 stathmin 1	3925	P16949
66	KRAS	MOCS2	MOCS2 molybdenum cofactor synthesis 2	4338	O96033
98	STK11	MOCS2	MOCS2 molybdenum cofactor synthesis 2	4338	O96033
67	KRAS	MT-CO3	MT-CO3 mitochondrially encoded cytochrome c oxidase III	4514	P00414
68	KRAS	MT-CYB	MT-CYB mitochondrially encoded cytochrome b	4519	P00156
97	STK11	MT-ND2	MT-ND2 NADH-ubiquinone oxidoreductase chain 2	4536	P03891
26	ATM	MT-ND4	MT-ND4 mitochondrially encoded NADH dehydrogenase 4	4538	P03905
37	CASP8	YBX1	YBX1 Y box binding protein 1	4904	P67809
9	AKT1	PDHB	PDHB pyruvate dehydrogenase (lipoamide) beta	5162	P11177

Appendix 4. List of interactions associated to BC detected by yeast two-hybrid library screens

11	AKT1	PLP1	PLP1 proteolipid protein 1	5354	P60201
55	ERBB2	POLD2	POLD2 polymerase (DNA directed), delta 2, regulatory subunit 50kDa	5425	P49005
12	AKT1	PRKCB	PRKCB protein kinase C, beta	5579	P05771
56	ERBB2	PSAP	PSAP prosaposin	5660	P07602
13	AKT1	PSMC5	PSMC5 proteasome (prosome, macropain) 26S subunit, ATPase, 5	5705	P62195
14	AKT1	PTGDS	PTGDS prostaglandin D2 synthase 21kDa (brain)	5730	P41222
15	AKT1	PTPN3	PTPN3 protein tyrosine phosphatase, non-receptor type 3	5774	P26045
34	CASP8	QDPR	QDPR quinoid dihydropteridine reductase	5860	P09417
69	KRAS	QDPR	QDPR quinoid dihydropteridine reductase	5860	P09417
16	AKT1	SNCB	SNCB synuclein, beta	6620	Q16143
35	CASP8	SOX2	SOX2 SRY (sex determining region Y)-box 2	6657	P48431
84	RB1CC1	SYN1	SYN1 synapsin I	6853	P17600
108	STK11	VPS72	VPS72 vacuolar protein sorting 72 homolog (S. cerevisiae)	6944	Q15906
104	STK11	THOP1	THOP1 thimet oligopeptidase 1	7064	P52888
45	CHEK2	TLR3	TLR3 toll-like receptor 3	7098	O15455
105	STK11	TSG101	TSG101 tumor susceptibility gene 101	7251	Q99816
19	AKT1	UCHL1	UCHL1 ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	7345	P09936
36	CASP8	UMPS	UMPS uridine monophosphate synthetase	7372	P11172
73	PIK3CA	UMPS	UMPS uridine monophosphate synthetase	7372	P11172
80	RAD51	UMPS	UMPS uridine monophosphate synthetase	7372	P11172
85	RB1CC1	UMPS	UMPS uridine monophosphate synthetase	7372	P11172
71	KRAS	UQCRC1	UQCRC1 ubiquinol-cytochrome c reductase core protein I	7384	P31930
32	CASP8	CNBP	CNBP CCHC-type zinc finger, nucleic acid binding protein	7555	P62633

106	STK11	TUBA1A	TUBA1A tubulin, alpha 1a	7846	Q71U36
57	ERBB2	STK24	STK24 serine/threonine kinase 24 (STE20 homolog, yeast)	8428	Q9Y6E0
103	STK11	STK24	STK24 serine/threonine kinase 24 (STE20 homolog, yeast)	8428	Q9Y6E0
83	RB1CC1	MADD	MADD MAP-kinase activating death domain	8567	Q8WXG6
70	KRAS	TNFSF13	TNFSF13 tumor necrosis factor (ligand) superfamily, member 13	8741	O75888
72	PIK3CA	TNFSF13	TNFSF13 tumor necrosis factor (ligand) superfamily, member 13	8741	O75888
81	RAD51	USP10	USP10 ubiquitin specific peptidase 10	9100	Q14694
22	ATM	C5orf13	C5orf13 chromosome 5 open reading frame 13 (PRO1873)	9315	Q16612
10	AKT1	PICK1	PICK1 protein interacting with PRKCA 1	9463	Q9NRD5
96	STK11	GNPDA1	GNPDA1 glucosamine-6-phosphate deaminase 1	10007	P46926
31	CASP8	BCL2L10	BCL2L10 BCL2-like 10 (apoptosis facilitator)	10017	Q9HD36
65	ESR1	UBE4B	UBE4B ubiquitination factor E4B (UFD2 homolog, yeast)	10277	O95155
107	STK11	TUBB4	TUBB4 tubulin, beta 4	10382	P04350
102	STK11	SEPT9	SEPT9 septin 9	10801	Q9UHD8
3	AKT1	CCNI	CCNI cyclin I	10983	Q14094
79	RAD51	SIRT2	SIRT2 sirtuin (silent mating type information regulation 2 homolog)	22933	Q8IXJ6
95	STK11	FAIM2	FAIM2 Fas apoptotic inhibitory molecule 2	23017	Q9BWQ8
4	AKT1	CLASP2	CLASP2 cytoplasmic linker associated protein 2	23122	O75122
77	RAD51	MAPK8IP3	MAPK8IP3 mitogen-activated protein kinase 8 interacting protein 3	23162	Q9UPT6
101	STK11	PLD3	PLD3 phospholipase D family, member 3	23646	Q8IV08
52	ERBB2	MTCH1	MTCH1 mitochondrial carrier homolog 1 (C. elegans)	23787	Q8IW90
18	AKT1	SULT4A1	SULT4A1 sulfotransferase family 4A, member 1	25830	Q9BR01
74	RAD51	COBRA1	COBRA1 cofactor of BRCA1	25920	Q8WX92
33	CASP8	HBP1	HBP1 HMG-box transcription factor 1	26959	O60381

Appendix 4. List of interactions associated to BC detected by yeast two-hybrid library screens

64	ESR1	HBP1	HBP1 HMG-box transcription factor 1	26959	O60381
82	RB1CC1	HBP1	HBP1 HMG-box transcription factor 1	26959	O60381
94	STK11	ERLEC1	ERLEC1 endoplasmic reticulum lectin 1	27248	Q96DZ1
78	RAD51	PCSK1N	PCSK1N proprotein convertase subtilisin/kexin type 1 inhibitor	27344	Q9UHG2
20	AKT1	ZNF691	ZNF691 zinc finger protein 691	51058	Q5VV52
86	STK11	ADIPOR1	ADIPOR1 adiponectin receptor 1	51094	Q96A54
61	ERBB2	UTP18	UTP18 UTP18, small subunit (SSU) processome component	51096	Q9Y5J1
49	ERBB2	GMPR2	GMPR2 guanosine monophosphate reductase 2	51292	Q9P2T1
21	ATM	ACTL6B	ACTL6B actin-like 6B	51412	O94805
2	AKT1	GET4	C7orf20 chromosome 7 open reading frame 20	51608	Q7L5D6
28	ATM	WHSC1L1	WHSC1L1 Wolf-Hirschhorn syndrome candidate 1-like 1	54904	Q9BZ95
25	ATM	MAP1S	MAP1S microtubule-associated protein 1S	55201	Q66K74
29	ATM	ZNF821	ZNF821 zinc finger protein 821	55565	O75541
99	STK11	PDXP	PDXP pyridoxal (pyridoxine, vitamin B6) phosphatase	57026	Q96GD0
17	AKT1	SRR	SRR serine racemase	63826	Q9GZT4
54	ERBB2	PHF23	PHF23 PHD finger protein 23	79142	Q9BUL5
100	STK11	PHF23	PHF23 PHD finger protein 23	79142	Q9BUL5
38	CHEK2	BAALC	BAALC brain and acute leukemia, cytoplasmic	79870	Q8WXS3
53	ERBB2	MUCL1	MUCL1 mucin-like 1	118430	Q96DR8
58	ERBB2	TCEAL2	TCEAL2 transcription elongation factor A (SII)-like 2	140597	Q9H3H9
60	ERBB2	TTC9B	TTC9B tetratricopeptide repeat domain 9B	148014	Q8N6N2
92	STK11	D2HGDH	D2HGDH D-2-hydroxyglutarate dehydrogenase	728294	Q8N465

Appendix 5. Functional modules detection in the BC-PIN

Homogeneous and/or enriched modules which include genes from our Y2H experiments are showed. For each module the following information is reported: a unique identifier (id); the number of proteins in the module (size); the driver(s) present in the module; the HC candidate(s) present in the module; functionally homogeneous modules (H); modules enriched for at least one GO biological process annotation (E); name of the enriched GO term(s).

id	size	Drivers	HC interactors	H	E	Enriched Term(s)
2	39	ERBB2	MUCL1, CLCN6, GMPR2, UTP18, TTC9B	y	y	signal transduction
6	27	AKT1	SRR, ACAT2, SULT4A1	y	y	protein amino acid phosphorylation
8	25	ATM, NBN, RAD50	C5orf13, MT-ND4	y	y	DNA repair
9	19	BCAR3, CASP8	CNBP, ATAD2	y	y	regulation of apoptosis
33	10	KRAS	MT-CYB, MOCS2, MT-CO3	y	y	Ras protein signal transduction
57	6	CHEK2	GNAS	y	y	signal transduction
71	5	FGFR2	RASL10B, PDLIM2	y	n	fibroblast growth factor receptor signaling pathway
93	4	XRCC3	SERPINB5, RNF20	y	n	DNA recombination;DNA repair
95	4	PTEN	ATP6AP1	y	n	cell death;angiogenesis;aging;cell migration

Appendix 5. Functional modules detection in the BC-PIN

166	3	FGFR4	WIF1	y	n	positive regulation of cell proliferation;fibroblast growth factor receptor signaling pathway
0	87	ESR2, ESR1		y	y	regulation of transcription, DNA-dependent;regulation of transcription
1	79	TP53		y	y	regulation of transcription
3	30	BRCA1		y	y	RNA splicing
4	30	SMAD4		y	y	regulation of transcription, DNA-dependent
5	29		CALM1	y	y	ion transport
7	25	TSG101		y	y	protein transport
10	17	TGFB1		y	y	cell adhesion
11	17		HSPA8	y	y	protein targeting to mitochondrion
12	17	RAF1		y	y	intracellular signaling cascade;Ras protein signal transduction
13	17	MAPK14		y	y	protein amino acid dephosphorylation
14	15		HSP90AA1	y	y	protein folding
15	15	HRAS		y	y	signal transduction
16	15		TUBB4	y	y	regulation of transcription
17	15	E2F1		y	y	regulation of transcription, DNA-dependent
18	14	PARP1		y	y	DNA repair
19	14	CDH1		y	y	positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle;negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle;anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process
20	13	AR		y	y	regulation of transcription, DNA-dependent
25	11		PRKCB	y	y	intracellular signaling cascade;protein amino acid phosphorylation
26	11		CCL5	y	n	G-protein coupled receptor protein signaling pathway
28	11		CFL2	y	y	protein folding
30	10		PSMC5	y	n	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle;negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle;anaphase-promoting complex-dependent proteasomal ubiquitin-dependent

						protein catabolic process
32	10	TGFBR1		y	y	protein amino acid phosphorylation
35	9		MTA3	y	y	regulation of transcription
36	9	MAP3K1		y	y	signal transduction
41	7		PICK1	y	y	regulation of transcription
42	7		CFL1	y	y	protein amino acid phosphorylation
48	7	BRMS1		y	y	negative regulation of transcription, DNA-dependent;transcription from RNA polymerase II promoter
51	6		SEPT9	y	n	cell cycle
54	6		BAI1	y	y	cell adhesion;G-protein coupled receptor protein signaling pathway;signal transduction;axonogenesis
56	6		GNPDA1, FASN, VPS72	y	y	regulation of transcription
61	5		USP10	y	y	ion transport;response to stimulus;sodium ion transport
62	5		CYP17A1	y	y	electron transport chain;oxidation reduction;transport
64	5	IGF1R		y	y	positive regulation of cell proliferation;positive regulation of cell migration;carbohydrate metabolic process;anti-apoptosis;aging;protein amino acid autophosphorylation
65	5		UQCRC1	y	n	electron transport chain;transport
67	5	RAD51		y	n	reciprocal meiotic recombination;double-strand break repair via homologous recombination
73	4	APC		y	y	intracellular signaling cascade
76	4		PDHB	y	n	pyruvate metabolic process;glycolysis
80	4		CXCL1	y	n	G-protein coupled receptor protein signaling pathway;signal transduction;inflammatory response;chemotaxis
82	4		TUBA1A	y	y	synaptic transmission
84	4		PLP1	y	y	synaptic transmission
90	4	AURKA		y	n	cell cycle

Appendix 5. Functional modules detection in the BC-PIN

94	4		BAALC	y	n	cell differentiation
97	4		POLD2	y	n	nucleotide-excision repair, DNA gap filling;DNA replication
98	4	BRCA2		y	n	regulation of transcription, DNA-dependent
99	4		CASC3	y	n	mRNA transport;transport;nuclear-transcribed mRNA catabolic process, nonsense-mediated decay;RNA splicing;mRNA processing
103	4		YBX1	y	n	RNA splicing;mRNA processing
104	4		COBRA1	y	y	negative regulation of transcription;regulation of transcription
105	4		CRY2	y	n	circadian rhythm
109	4		CSRP1	y	n	negative regulation of cellular component movement;carbohydrate metabolic process;regulation of apoptosis;focal adhesion assembly;cellular response to insulin stimulus;fructose metabolic process;fructose 6-phosphate metabolic process;gluconeogenesis
110	4	TGFBR3		y	n	BMP signaling pathway
111	4		SOX2	y	y	regulation of transcription, DNA-dependent;positive regulation of transcription from RNA polymerase II promoter
113	4		PTGDS	y	n	G-protein coupled receptor protein signaling pathway;signal transduction
116	3		MAPK8IP3	y	n	regulation of JNK cascade;protein amino acid phosphorylation;activation of JUN kinase activity
120	3		SLC25A6	y	n	apoptosis
125	3	PAX2		y	n	visual perception
128	3	PGR		y	n	progesterone receptor signaling pathway;regulation of transcription, DNA-dependent
136	3	CDKN2A		y	n	DNA replication;cell cycle
138	3		SNCB	y	y	response to drug;signal transduction
146	3		PCSK1N	y	n	proteolysis;peptide hormone processing
148	3		HMGN1	y	n	regulation of transcription, DNA-dependent;protein amino acid phosphorylation
153	3	BRIP1		y	n	DNA damage response, signal transduction resulting in induction of apoptosis
155	3		MAP1S	y	n	mitochondrion transport along microtubule
157	3	PTPN1		y	n	signal transduction

158	3	SNAI1	y	y	multicellular organismal development
159	3	STK24	y	n	protein amino acid phosphorylation
162	3	ZNF691	y	n	embryonic organ morphogenesis;endocrine pancreas development;glucose homeostasis;multicellular organismal development;insulin secretion;vesicle transport along microtubule;positive regulation of neuron differentiation;regulation of insulin secretion;synaptic transmission;negative regulation of apoptosis;neurogenesis;cerebellum development;nitric oxide mediated signal transduction;anterior/posterior pattern formation;cell fate commitment;brain development;nervous system development;response to drug;response to glucose stimulus;regulation of transcription, DNA-dependent;positive regulation of transcription from RNA polymerase II promoter;hindbrain development;camera-type eye development
167	3	PHF23	y	n	L-serine catabolic process;protein homotetramerization;glycine metabolic process;one-carbon metabolic process;protein tetramerization;purine base biosynthetic process;glycine biosynthetic process from serine;proteolysis;folic acid metabolic process
172	3	NR4A1	y	n	signal transduction

Appendix 6. List of CRC causative genes (drivers) and their features

Genes that have been also examined by Y2H library screens are highlighted in orange.

Symbol	Gene name	UniProt	Expression pattern in CRC tissue
APC	Adenomatous polyposis coli protein	P25054	moderate down
AURKA	Serine/threonine-protein kinase 6	O14965	up
AXIN2	Axin-2	Q9Y2T1	no
BAX	Apoptosis regulator BAX, cytoplasmic isoform beta	Q07814	moderate up
BCL10	B-cell lymphoma/leukemia 10	O95999	down
BMPR1A	Bone morphogenetic protein receptor type IA precursor	P36894	up
BRAF	B-Raf proto-oncogene serine/threonine-protein kinase	P15056	moderate down
BUB1	Mitotic checkpoint serine/threonine-protein kinase BUB1	O43683	up
BUB1B	Mitotic checkpoint serine/threonine-protein kinase BUB1 beta	O60566	up
CCND1	G1/S-specific cyclin-D1	P24385	up
CDH1	Epithelial-cadherin precursor	P12830	moderate down
CDKN2A	Cyclin-dependent kinase inhibitor 2A, isoform 4	Q8N726	n/a
CTNNA1	Catenin alpha-1	P35221	moderate down
CTNNB1	Catenin beta-1	P35222	up
DCC	Netrin receptor DCC precursor	P43146	n/a
DLC1	Rho GTPase-activating protein 7	Q96QB1	up
EGFR	Epidermal growth factor receptor precursor	P00533	down
ERBB2	Receptor tyrosine-protein kinase erbB-2 precursor	P04626	moderate down
FBXW7	F-box/WD repeat protein 7	Q969H0	moderate up
FLCN	Folliculin	Q8NFG4	moderate down

KRAS	GTPase KRas precursor	P01116	down
MAP2K4	Dual specificity mitogen-activated protein kinase kinase 4	P45985	moderate down
MCC	Colorectal mutant cancer protein	P23508	n/a
MLH1	DNA mismatch repair protein Mlh1	P40692	no
MLH3	DNA mismatch repair protein Mlh3	Q9UHC1	no
MSH2	DNA mismatch repair protein Msh2	P43246	up
MSH6	DNA mismatch repair protein MSH6	P52701	up
MUTYH	A/G-specific adenine DNA glycosylase	Q9UIF7	no
NRAS	GTPase NRas precursor	P01111	moderate down
ODC1	Ornithine decarboxylase	P11926	up
PDGFRL	Platelet-derived growth factor receptor-like protein precursor	Q15198	up
PIK3CA	PIP2 3-kinase catalytic subunit alpha isoform	P42336	moderate down
PMS2	PMS1 protein homolog 2	P54278	no
PTEN	PIP3 3-phosphatase and dual- specificity protein phosphatase	P60484	down
PTPN12	Tyrosine-protein phosphatase non-receptor type 12	Q05209	up
PTPRJ	Receptor-type tyrosine-protein phosphatase eta precursor	Q12913	n/a
RAD54B	DNA repair and recombination protein RAD54B	Q9Y620	up
RB1	Retinoblastoma-associated protein	P06400	up
SMAD2	Mothers against decapentaplegic homolog 2	Q15796	moderate down
SMAD4	Mothers against decapentaplegic homolog 4	Q13485	moderate down
SRC	Proto-oncogene tyrosine-protein kinase Src	P12931	n/a
STK11	Serine/threonine-protein kinase 11	Q15831	moderate down
TGFBR2	TGF-beta receptor type-2 precursor	P37173	no
TLR2	Toll-like receptor 2 precursor	O60603	up
TP53	Cellular tumor antigen p53	P04637	no

Appendix 7. List of CRC candidate genes and their features

Genes that have been also examined by Y2H library screens are highlighted in orange.

Symbol	Gene name	UniProt	Expression pattern in CRC tissue
AKT1	RAC-alpha serine/threonine-protein kinase	P31749	up
ALDOB	Fructose-bisphosphate aldolase B	P05062	no
ANP32B	Acidic leucine-rich nuclear phosphoprotein 32 family member B	Q92688	up
BAAT	Bile acid CoA:amino acid N-acyltransferase	Q14032	n/a
C9orf156	Nef-associated protein 1	Q9BU70	moderate down
C9orf3	Aminopeptidase O	Q8N6M6	moderate down
C9orf30	Uncharacterized protein C9orf30	Q96H12	up
C9orf97	Uncharacterized protein C9orf97	Q5T7W7	moderate down
CDC14B	Dual specificity protein phosphatase CDC14B	O60729	moderate down
CDK8	Cell division protein kinase 8	P49336	up
CORO2A	Coronin-2A	Q92828	down
CTSL2	Cathepsin L2 precursor	O60911	moderate up
CYLC2	Cylicin-2	Q14093	n/a
DVL1	Segment polarity protein dishevelled homolog DVL-1	O14640	up
FANCC	Fanconi anemia group C protein	Q00597	n/a
FBP1	Fructose-1,6-bisphosphatase 1	P09467	no
FBP2	Fructose-1,6-bisphosphatase isozyme 2	O00757	moderate up
FRAT2	GSK-3-binding protein FRAT2	O75474	down
GALNT12	Polypeptide N-acetylgalactosaminyltransferase 12	Q8IXK2	down
HEMGN	Hemogen	Q9BXL5	n/a

HRAS	GTPase HRas precursor	P01112	up
HSD17B3	Estradiol 17-beta-dehydrogenase 3	P37058	n/a
IGFBP3	Insulin-like growth factor-binding protein 3 precursor	P17936	up
KIAA1529	Protein KIAA1529 Fragment	Q68DP5	n/a
LEF1	Lymphoid enhancer-binding factor 1	Q9UJU2	up
MAP2K1	Dual specificity mitogen-activated protein kinase kinase 1	Q02750	up
MAPK3	Mitogen-activated protein kinase 3	Q16644	down
NANS	Sialic acid synthase	Q9NR45	down
PPP2CB	Ser/thr-protein phosphatase 2A catalytic subunit beta isoform	P62714	down
PPP3R2	Calcineurin subunit B isoform 2	Q96LZ3	n/a
RASA1	Ras GTPase-activating protein 1	P20936	down
SEC61B	Protein transport protein Sec61 subunit beta	P60468	up
SFRP2	Secreted frizzled-related protein 2 precursor	Q96HF1	up
SFRP4	Secreted frizzled-related protein 4 precursor	Q6FHJ7	up
SHC1	SHC-transforming protein 1	P29353	up
SMAD1	Mothers against decapentaplegic homolog 1	Q15797	down
STX17	Syntaxin-17	P56962	moderate down
TBC1D2	TBC1 domain family member 2	Q9BYX2	no
TDRD7	Tudor domain-containing protein 7	Q8NHU6	moderate down
TGFB1	Transforming growth factor beta-1 precursor	P01137	up
TMEFF1	Tomoregulin-1 precursor	Q8IYR6	up
TMOD1	Tropomodulin-1	P28289	n/a
XPA	DNA-repair protein complementing XP-A cells	P23025	moderate down
ZNF189	Zinc finger protein 189	O75820	moderate up
ZNF510	Zinc finger protein 510	Q9Y2H8	moderate dow

Appendix 8. List of interactions associated to CRC detected by Y2H matrix screens

Interacting pairs are reported as gene symbols taken from HGNC. The confidence level (high or low) is reported.

HIGH CONFIDENCE

	Bait	Prey
1	APC	ANP32B
2	APC	BAAT
3	APC	C9orf97
4	APC	CTSL2
5	APC	FANCC
6	APC	FBP1
7	APC	MAP2K1
8	APC	NANS
9	APC	PPP3R2
10	APC	RASA1
11	APC	SMAD1
12	APC	TGFB1
13	APC	TMEFF1
14	APC	TMOD1
15	APC	ZNF510
16	AURKA	AKT1
17	AURKA	ANP32B
18	AURKA	BAAT
19	AURKA	C9orf156
20	AURKA	C9orf97
21	AURKA	CDK8
22	AURKA	CORO2A
23	AURKA	CYLC2
24	AURKA	FBP2
25	AURKA	HEMGN
26	AURKA	IGFBP3
27	AURKA	LEF1
28	AURKA	MAP2K1
29	AURKA	MAPK3
30	AURKA	NANS
31	AURKA	PPP2CB
32	AURKA	PPP3R2
33	AURKA	RASA1
34	AURKA	SEC61B
35	AURKA	SFRP4
36	AURKA	STX17
37	AURKA	TBC1D2
38	AURKA	TGFB1
39	AURKA	XPA
40	AURKA	ZNF189
41	AURKA	ZNF510
42	AXIN2	FBP1
43	AXIN2	FBP2
44	AXIN2	HEMGN
45	AXIN2	SMAD1
46	AXIN2	TGFB1
47	AXIN2	ZNF189
48	BAX	ANP32B
49	BAX	LEF1
50	BAX	MAP2K1
51	BCL10	AKT1
52	BCL10	ANP32B
53	BCL10	BAAT
54	BCL10	C9orf156
55	BCL10	CDC14B
56	BCL10	FBP1
57	BCL10	HEMGN
58	BCL10	KIAA1529
59	BCL10	NANS
60	BCL10	RASA1
61	BCL10	TGFB1
62	BMPR1A	ALDOB
63	BMPR1A	C9orf156
64	BMPR1A	C9orf3
65	BMPR1A	C9orf30
66	BMPR1A	CDK8
67	BMPR1A	CTSL2
68	BMPR1A	FANCC
69	BMPR1A	FBP1
70	BMPR1A	FBP2
71	BMPR1A	GALNT12
72	BMPR1A	HRAS
73	BMPR1A	IGFBP3
74	BMPR1A	MAP2K1
75	BMPR1A	RASA1
76	BMPR1A	SEC61B
77	BMPR1A	SFRP2
78	BMPR1A	SFRP4
79	BMPR1A	SMAD1
80	BMPR1A	TGFB1
81	BMPR1A	XPA
82	BMPR1A	ZNF189
83	BRAF	CTSL2
84	BRAF	FBP1

85	BRAF	HSD17B3
86	BRAF	MAPK3
87	BRAF	ZNF189
88	BRAF	ZNF510
89	BUB1	AKT1
90	BUB1	ALDOB
91	BUB1	ANP32B
92	BUB1	BAAT
93	BUB1	C9orf156
94	BUB1	C9orf3
95	BUB1	C9orf30
96	BUB1	C9orf97
97	BUB1	CDC14B
98	BUB1	CDK8
99	BUB1	CORO2A
100	BUB1	CTSL2
101	BUB1	CYLC2
102	BUB1	DVL1
103	BUB1	FANCC
104	BUB1	FBP1
105	BUB1	FBP2
106	BUB1	FRAT2
107	BUB1	GALNT12
108	BUB1	HEMGN
109	BUB1	HRAS
110	BUB1	HSD17B3
111	BUB1	IGFBP3
112	BUB1	KIAA1529
113	BUB1	LEF1
114	BUB1	MAP2K1

115	BUB1	MAPK3
116	BUB1	NANS
117	BUB1	PPP2CB
118	BUB1	PPP3R2
119	BUB1	RASA1
120	BUB1	SEC61B
121	BUB1	SFRP2
122	BUB1	SFRP4
123	BUB1	SHC1
124	BUB1	SMAD1
125	BUB1	STX17
126	BUB1	TBC1D2
127	BUB1	TDRD7
128	BUB1	TGFB1
129	BUB1	TMEFF1
130	BUB1	TMOD1
131	BUB1	XPA
132	BUB1	ZNF189
133	BUB1	ZNF510
134	CCND1	C9orf156
135	CCND1	C9orf97
136	CCND1	CDC14B
137	CCND1	CDK8
138	CCND1	FANCC
139	CCND1	IGFBP3
140	CCND1	PPP3R2
141	CCND1	SMAD1
142	CCND1	TBC1D2
143	CCND1	TDRD7
144	CCND1	ZNF510

145	CDH1	AKT1
146	CDH1	C9orf3
147	CDH1	C9orf30
148	CDH1	CDK8
149	CDH1	GALNT12
150	CDH1	HEMGN
151	CDH1	HRAS
152	CDH1	HSD17B3
153	CDH1	MAP2K1
154	CDH1	MAPK3
155	CDH1	NANS
156	CDH1	SFRP2
157	CDH1	STX17
158	CDH1	TMOD1
159	CDH1	ZNF510
160	CDKN2A	ALDOB
161	CDKN2A	BAAT
162	CDKN2A	C9orf3
163	CDKN2A	CDC14B
164	CDKN2A	CTSL2
165	CDKN2A	CYLC2
166	CDKN2A	FRAT2
167	CDKN2A	GALNT12
168	CDKN2A	HRAS
169	CDKN2A	KIAA1529
170	CDKN2A	MAP2K1
171	CDKN2A	MAPK3
172	CDKN2A	RASA1
173	CDKN2A	SEC61B
174	CDKN2A	SFRP4

175	CDKN2A	TGFB1
176	CDKN2A	TMEFF1
177	CDKN2A	TMOD1
178	CTNNA1	AKT1
179	CTNNA1	ALDOB
180	CTNNA1	BAAT
181	CTNNA1	CTSL2
182	CTNNA1	FBP1
183	CTNNA1	FBP2
184	CTNNA1	FRAT2
185	CTNNA1	HRAS
186	CTNNA1	HSD17B3
187	CTNNA1	KIAA1529
188	CTNNA1	MAP2K1
189	CTNNA1	NANS
190	CTNNA1	SFRP2
191	CTNNA1	SFRP4
192	CTNNA1	SMAD1
193	CTNNA1	STX17
194	CTNNA1	TDRD7
195	CTNNA1	ZNF189
196	CTNNA1	ZNF510
197	DCC	ALDOB
198	DCC	ANP32B
199	DCC	C9orf156
200	DCC	C9orf97
201	DCC	CORO2A
202	DCC	CTSL2
203	DCC	FBP1
204	DCC	HEMGN

Appendix 8. List of interactions associated to CRC detected by Y2H matrix screens

205	DCC	MAPK3
206	DCC	RASA1
207	DCC	SFRP2
208	DCC	TBC1D2
209	DCC	TDRD7
210	DCC	TMOD1
211	DLC1	BAAT
212	DLC1	C9orf156
213	DLC1	C9orf3
214	DLC1	CDC14B
215	DLC1	CORO2A
216	DLC1	CTSL2
217	DLC1	FANCC
218	DLC1	FBP2
219	DLC1	HEMGN
220	DLC1	HSD17B3
221	DLC1	KIAA1529
222	DLC1	MAPK3
223	DLC1	PPP3R2
224	DLC1	RASA1
225	DLC1	SFRP4
226	DLC1	SMAD1
227	DLC1	STX17
228	DLC1	TDRD7
229	DLC1	TGFB1
230	DLC1	TMEFF1
231	DLC1	TMOD1
232	DLC1	XPA
233	DLC1	ZNF189
234	DLC1	ZNF510
235	EGFR	C9orf156
236	EGFR	C9orf3
237	EGFR	C9orf97
238	EGFR	CYLC2
239	EGFR	FANCC
240	EGFR	FRAT2
241	EGFR	HEMGN
242	EGFR	MAPK3
243	EGFR	NANS
244	EGFR	PPP2CB
245	EGFR	STX17
246	EGFR	TGFB1
247	EGFR	ZNF510
248	ERBB2	BAAT
249	ERBB2	C9orf156
250	ERBB2	CDK8
251	ERBB2	CYLC2
252	ERBB2	FBP1
253	ERBB2	FRAT2
254	ERBB2	HSD17B3
255	ERBB2	KIAA1529
256	ERBB2	MAP2K1
257	ERBB2	RASA1
258	ERBB2	SFRP4
259	ERBB2	SMAD1
260	ERBB2	TGFB1
261	ERBB2	ZNF189
262	FBXW7	AKT1
263	FBXW7	ANP32B
264	FBXW7	DVL1
265	FBXW7	FANCC
266	FBXW7	FBP1
267	FBXW7	FBP2
268	FBXW7	GALNT12
269	FBXW7	HEMGN
270	FBXW7	HRAS
271	FBXW7	IGFBP3
272	FBXW7	MAP2K1
273	FBXW7	MAPK3
274	FBXW7	NANS
275	FBXW7	PPP3R2
276	FBXW7	SEC61B
277	FBXW7	SHC1
278	FBXW7	SMAD1
279	FBXW7	TGFB1
280	FBXW7	TMOD1
281	FBXW7	XPA
282	FBXW7	ZNF510
283	FLCN	ALDOB
284	FLCN	C9orf97
285	FLCN	FBP1
286	FLCN	HEMGN
287	FLCN	KIAA1529
288	FLCN	NANS
289	FLCN	TMEFF1
290	KRAS	CDK8
291	KRAS	FANCC
292	KRAS	FRAT2
293	KRAS	HEMGN
294	KRAS	IGFBP3
295	KRAS	PPP2CB
296	KRAS	STX17
297	KRAS	ZNF189
298	MAP2K4	ALDOB
299	MAP2K4	CYLC2
300	MAP2K4	FBP2
301	MAP2K4	STX17
302	MCC	ALDOB
303	MCC	ANP32B
304	MCC	BAAT
305	MCC	C9orf156
306	MCC	C9orf30
307	MCC	C9orf97
308	MCC	CDC14B
309	MCC	CDK8
310	MCC	CYLC2
311	MCC	FRAT2
312	MCC	HEMGN
313	MCC	HSD17B3
314	MCC	KIAA1529
315	MCC	NANS
316	MCC	SEC61B
317	MCC	SFRP4
318	MCC	TGFB1
319	MCC	ZNF189
320	MLH1	ALDOB
321	MLH1	BAAT
322	MLH1	C9orf156
323	MLH1	CTSL2
324	MLH1	CYLC2

325	MLH1	FRAT2
326	MLH1	KIAA1529
327	MLH1	LEF1
328	MLH1	PPP2CB
329	MLH1	SMAD1
330	MLH1	STX17
331	MLH1	TDRD7
332	MLH1	XPA
333	MLH3	AKT1
334	MLH3	ALDOB
335	MLH3	ANP32B
336	MLH3	BAAT
337	MLH3	C9orf3
338	MLH3	C9orf30
339	MLH3	C9orf97
340	MLH3	CDC14B
341	MLH3	CDK8
342	MLH3	CTSL2
343	MLH3	CYLC2
344	MLH3	DVL1
345	MLH3	FANCC
346	MLH3	FBP1
347	MLH3	FBP2
348	MLH3	FRAT2
349	MLH3	GALNT12
350	MLH3	HEMGN
351	MLH3	HRAS
352	MLH3	HSD17B3
353	MLH3	IGFBP3
354	MLH3	KIAA1529

355	MLH3	LEF1
356	MLH3	MAP2K1
357	MLH3	NANS
358	MLH3	PPP2CB
359	MLH3	PPP3R2
360	MLH3	RASA1
361	MLH3	SEC61B
362	MLH3	SFRP2
363	MLH3	SMAD1
364	MLH3	STX17
365	MLH3	TBC1D2
366	MLH3	TDRD7
367	MLH3	TGFB1
368	MLH3	TMEFF1
369	MLH3	TMOD1
370	MLH3	XPA
371	MLH3	ZNF510
372	MSH2	AKT1
373	MSH2	CDC14B
374	MSH2	DVL1
375	MSH2	FBP1
376	MSH2	FBP2
377	MSH2	GALNT12
378	MSH2	HRAS
379	MSH2	KIAA1529
380	MSH2	LEF1
381	MSH2	PPP3R2
382	MSH2	SMAD1
383	MSH2	STX17
384	MSH2	TDRD7

385	MSH2	ZNF510
386	MSH6	HRAS
387	MSH6	MAP2K1
388	MSH6	NANS
389	MUTYH	BAAT
390	MUTYH	C9orf30
391	MUTYH	C9orf97
392	MUTYH	CYLC2
393	MUTYH	FANCC
394	MUTYH	HRAS
395	MUTYH	SEC61B
396	MUTYH	SMAD1
397	MUTYH	TDRD7
398	MUTYH	XPA
399	NRAS	AKT1
400	NRAS	ALDOB
401	NRAS	C9orf156
402	NRAS	C9orf3
403	NRAS	CORO2A
404	NRAS	CYLC2
405	NRAS	FANCC
406	NRAS	FBP2
407	NRAS	FRAT2
408	NRAS	HEMGN
409	NRAS	KIAA1529
410	NRAS	LEF1
411	NRAS	MAPK3
412	NRAS	PPP2CB
413	NRAS	RASA1
414	NRAS	SFRP4

415	NRAS	SMAD1
416	NRAS	STX17
417	NRAS	TDRD7
418	NRAS	XPA
419	ODC1	BAAT
420	ODC1	C9orf97
421	ODC1	HEMGN
422	ODC1	MAP2K1
423	ODC1	NANS
424	ODC1	STX17
425	ODC1	TDRD7
426	ODC1	TMOD1
427	ODC1	ZNF510
428	PDGFRL	BAAT
429	PDGFRL	CTSL2
430	PDGFRL	FANCC
431	PDGFRL	FBP2
432	PDGFRL	HRAS
433	PDGFRL	KIAA1529
434	PDGFRL	MAP2K1
435	PDGFRL	MAPK3
436	PDGFRL	NANS
437	PDGFRL	PPP2CB
438	PDGFRL	PPP3R2
439	PDGFRL	SFRP4
440	PDGFRL	SMAD1
441	PDGFRL	TBC1D2
442	PDGFRL	TGFB1
443	PDGFRL	XPA
444	PDGFRL	ZNF510

Appendix 8. List of interactions associated to CRC detected by Y2H matrix screens

445	PIK3CA	FANCC
446	PIK3CA	FBP2
447	PIK3CA	GALNT12
448	PIK3CA	HRAS
449	PIK3CA	MAP2K1
450	PIK3CA	SFRP4
451	PIK3CA	TMOD1
452	PMS2	AKT1
453	PMS2	ANP32B
454	PMS2	C9orf156
455	PMS2	FANCC
456	PMS2	GALNT12
457	PMS2	KIAA1529
458	PMS2	PPP2CB
459	PMS2	RASA1
460	PMS2	SFRP4
461	PMS2	XPA
462	PMS2	ZNF189
463	PTEN	ANP32B
464	PTEN	C9orf156
465	PTEN	KIAA1529
466	PTEN	TMEFF1
467	PTPN12	ANP32B
468	PTPN12	FANCC
469	PTPN12	FRAT2
470	PTPN12	KIAA1529
471	PTPN12	PPP3R2
472	PTPN12	SFRP4
473	PTPN12	SHC1
474	PTPN12	SMAD1

475	PTPN12	STX17
476	PTPN12	ZNF510
477	PTPRJ	ANP32B
478	PTPRJ	BAAT
479	PTPRJ	C9orf156
480	PTPRJ	C9orf97
481	PTPRJ	FANCC
482	PTPRJ	FBP1
483	PTPRJ	HEMGN
484	PTPRJ	HRAS
485	PTPRJ	MAP2K1
486	PTPRJ	NANS
487	PTPRJ	PPP2CB
488	PTPRJ	PPP3R2
489	PTPRJ	RASA1
490	PTPRJ	SFRP4
491	PTPRJ	TMEFF1
492	PTPRJ	ZNF510
493	RAD54B	BAAT
494	RAD54B	C9orf156
495	RAD54B	IGFBP3
496	RAD54B	SFRP2
497	RB1	BAAT
498	RB1	C9orf156
499	RB1	CORO2A
500	RB1	CTSL2
501	RB1	DVL1
502	RB1	FANCC
503	RB1	FBP1
504	RB1	FBP2

505	RB1	GALNT12
506	RB1	KIAA1529
507	RB1	LEF1
508	RB1	MAPK3
509	RB1	RASA1
510	RB1	SHC1
511	RB1	STX17
512	RB1	XPA
513	SMAD2	AKT1
514	SMAD2	C9orf156
515	SMAD2	CORO2A
516	SMAD2	CYLC2
517	SMAD2	FBP2
518	SMAD2	HSD17B3
519	SMAD2	MAP2K1
520	SMAD2	RASA1
521	SMAD2	XPA
522	SMAD2	ZNF510
523	SMAD4	ALDOB
524	SMAD4	BAAT
525	SMAD4	C9orf3
526	SMAD4	C9orf30
527	SMAD4	C9orf97
528	SMAD4	CDC14B
529	SMAD4	FANCC
530	SMAD4	FBP2
531	SMAD4	HEMGN
532	SMAD4	HRAS
533	SMAD4	KIAA1529
534	SMAD4	PPP2CB

535	SMAD4	PPP3R2
536	SMAD4	SFRP2
537	SMAD4	SFRP4
538	SMAD4	SHC1
539	SMAD4	SMAD1
540	SMAD4	STX17
541	SMAD4	ZNF189
542	SMAD4	ZNF510
543	SRC	ALDOB
544	SRC	BAAT
545	SRC	C9orf156
546	SRC	CTSL2
547	SRC	FANCC
548	SRC	FBP2
549	SRC	GALNT12
550	SRC	HEMGN
551	SRC	KIAA1529
552	SRC	MAP2K1
553	SRC	MAPK3
554	SRC	NANS
555	SRC	PPP2CB
556	SRC	STX17
557	SRC	XPA
558	SRC	ZNF189
559	STK11	AKT1
560	STK11	ALDOB
561	STK11	ANP32B
562	STK11	C9orf156
563	STK11	CDC14B
564	STK11	CYLC2

565	STK11	FBP1
566	STK11	FBP2
567	STK11	HRAS
568	STK11	IGFBP3
569	STK11	KIAA1529
570	STK11	LEF1
571	STK11	MAP2K1
572	STK11	MAPK3
573	STK11	NANS
574	STK11	PPP2CB
575	STK11	RASA1
576	STK11	STX17
577	STK11	TDRD7
578	STK11	XPA
579	STK11	ZNF189
580	STK11	ZNF510
581	TGFBR2	C9orf156
582	TGFBR2	FANCC
583	TGFBR2	MAP2K1
584	TGFBR2	PPP2CB
585	TGFBR2	ZNF510
586	TLR2	AKT1
587	TLR2	ANP32B
588	TLR2	FBP1
589	TLR2	HEMGN
590	TLR2	MAP2K1
591	TLR2	NANS
592	TLR2	SMAD1
593	TLR2	STX17
594	TLR2	TGFB1
595	TLR2	TMOD1

LOW CONFIDENCE

	Bait	Prey
596	APC	FRAT2
597	APC	HEMGN
598	APC	HSD17B3
599	APC	MAPK3
600	APC	XPA
601	AURKA	ALDOB
602	AURKA	C9orf30
603	AURKA	CDC14B
604	AURKA	FBP1
605	AURKA	FRAT2
606	AURKA	HRAS
607	AURKA	HSD17B3
608	AURKA	KIAA1529
609	AURKA	SMAD1
610	AURKA	TDRD7
611	AURKA	TMEFF1
612	AURKA	TMOD1
613	AXIN2	AKT1
614	AXIN2	ALDOB
615	AXIN2	BAAT
616	AXIN2	C9orf156
617	AXIN2	C9orf97
618	AXIN2	CDC14B
619	AXIN2	CYLC2
620	AXIN2	FRAT2
621	AXIN2	PPP3R2
622	BAX	CDC14B
623	BAX	DVL1
624	BAX	HEMGN
625	BAX	PPP2CB
626	BAX	TMOD1
627	BAX	XPA
628	BCL10	C9orf30
629	BCL10	C9orf97
630	BCL10	CDK8
631	BCL10	CORO2A
632	BCL10	FANCC
633	BCL10	HRAS
634	BCL10	HSD17B3
635	BCL10	MAP2K1
636	BCL10	MAPK3
637	BCL10	SEC61B
638	BCL10	SFRP4
639	BCL10	TMEFF1
640	BCL10	TMOD1
641	BMPR1A	ANP32B
642	BMPR1A	BAAT
643	BMPR1A	C9orf97
644	BMPR1A	CDC14B
645	BMPR1A	HEMGN
646	BMPR1A	KIAA1529
647	BMPR1A	LEF1
648	BMPR1A	MAPK3
649	BMPR1A	PPP2CB
650	BMPR1A	PPP3R2
651	BMPR1A	SHC1
652	BMPR1A	STX17
653	BMPR1A	TBC1D2
654	BMPR1A	TDRD7
655	BMPR1A	TMEFF1
656	BMPR1A	ZNF510
657	BRAF	BAAT
658	BRAF	C9orf3
659	BRAF	LEF1
660	CCND1	ALDOB
661	CCND1	ANP32B
662	CCND1	BAAT
663	CCND1	CORO2A
664	CCND1	CTSL2
665	CCND1	CYLC2
666	CCND1	DVL1
667	CCND1	FRAT2
668	CCND1	HEMGN
669	CCND1	HRAS
670	CCND1	HSD17B3
671	CCND1	KIAA1529
672	CCND1	LEF1
673	CCND1	PPP2CB
674	CCND1	SEC61B
675	CCND1	SFRP2
676	CCND1	SFRP4
677	CCND1	TGFB1
678	CCND1	TMEFF1
679	CCND1	TMOD1

Appendix 8. List of interactions associated to CRC detected by Y2H matrix screens

680	CCND1	ZNF189
681	CDH1	ANP32B
682	CDH1	CDC14B
683	CDH1	FANCC
684	CDH1	FBP1
685	CDH1	IGFBP3
686	CDH1	RASA1
687	CDH1	TGFB1
688	CDH1	ZNF189
689	CDKN2A	FBP2
690	CDKN2A	PPP3R2
691	CDKN2A	SHC1
692	CDKN2A	SMAD1
693	CDKN2A	STX17
694	CDKN2A	ZNF510
695	CTNNA1	C9orf156
696	CTNNA1	CDC14B
697	CTNNA1	CYLC2
698	CTNNA1	GALNT12
699	CTNNA1	HEMGN
700	CTNNA1	TMEFF1
701	CTNNA1	XPA
702	DCC	CDK8
703	DCC	DVL1
704	DCC	FANCC
705	DCC	FBP2
706	DCC	HSD17B3
707	DCC	IGFBP3
708	DCC	KIAA1529
709	DCC	MAP2K1
710	DCC	PPP2CB
711	DCC	PPP3R2
712	DCC	SEC61B
713	DCC	ZNF510
714	DLC1	AKT1
715	DLC1	ANP32B
716	DLC1	CDK8
717	DLC1	CYLC2
718	DLC1	DVL1
719	DLC1	FBP1
720	DLC1	HRAS
721	DLC1	LEF1
722	DLC1	NANS
723	DLC1	SEC61B
724	DLC1	SHC1
725	DLC1	TBC1D2
726	EGFR	ALDOB
727	EGFR	BAAT
728	EGFR	CDK8
729	EGFR	CORO2A
730	EGFR	CTSL2
731	EGFR	FBP2
732	EGFR	GALNT12
733	EGFR	HSD17B3
734	EGFR	MAP2K1
735	EGFR	RASA1
736	EGFR	SFRP2
737	EGFR	SFRP4
738	EGFR	SMAD1
739	EGFR	TBC1D2
740	EGFR	TDRD7
741	ERBB2	AKT1
742	ERBB2	ANP32B
743	ERBB2	FANCC
744	ERBB2	HRAS
745	ERBB2	NANS
746	ERBB2	SEC61B
747	ERBB2	STX17
748	ERBB2	TDRD7
749	ERBB2	TMOD1
750	ERBB2	ZNF510
751	FBXW7	FANCC
752	FBXW7	BAAT
753	FBXW7	C9orf156
754	FBXW7	C9orf30
755	FBXW7	C9orf97
756	FBXW7	CDC14B
757	FBXW7	CDK8
758	FBXW7	CTSL2
759	FBXW7	FRAT2
760	FBXW7	HSD17B3
761	FBXW7	KIAA1529
762	FBXW7	LEF1
763	FBXW7	RASA1
764	FBXW7	SFRP4
765	FBXW7	TBC1D2
766	FBXW7	TDRD7
767	FBXW7	TMEFF1
768	FBXW7	ZNF189
769	FLCN	AKT1
770	FLCN	ANP32B
771	FLCN	C9orf156
772	FLCN	C9orf3
773	FLCN	C9orf30
774	FLCN	CYLC2
775	FLCN	GALNT12
776	FLCN	HRAS
777	FLCN	HSD17B3
778	FLCN	MAP2K1
779	FLCN	RASA1
780	FLCN	SFRP2
781	FLCN	SHC1
782	FLCN	TGFB1
783	FLCN	TMOD1
784	FLCN	ZNF189
785	FLCN	ZNF510
786	KRAS	AKT1
787	KRAS	ALDOB
788	KRAS	ANP32B
789	KRAS	BAAT
790	KRAS	C9orf156
791	KRAS	C9orf97
792	KRAS	CORO2A
793	KRAS	CTSL2
794	KRAS	CYLC2
795	KRAS	FBP1
796	KRAS	FBP2
797	KRAS	GALNT12
798	KRAS	LEF1
799	KRAS	MAPK3

800	KRAS	SEC61B
801	KRAS	SHC1
802	KRAS	SMAD1
803	KRAS	TBC1D2
804	KRAS	TGFB1
805	KRAS	TMOD1
806	KRAS	XPA
807	MAP2K4	BAAT
808	MAP2K4	C9orf97
809	MAP2K4	FANCC
810	MAP2K4	KIAA1529
811	MAP2K4	PPP3R2
812	MAP2K4	SMAD1
813	MAP2K4	TDRD7
814	MAP2K4	ZNF510
815	MCC	C9orf3
816	MCC	GALNT12
817	MCC	LEF1
818	MCC	SFRP2
819	MCC	STX17
820	MLH1	FANCC
821	MLH1	FBP1
822	MLH1	FBP2
823	MLH1	NANS
824	MLH1	SFRP2
825	MLH1	TMOD1
826	MLH3	CORO2A
827	MLH3	SFRP4
828	MLH3	ZNF189
829	MSH2	BAAT

830	MSH2	C9orf30
831	MSH2	CTSL2
832	MSH2	CYLC2
833	MSH2	FANCC
834	MSH2	FRAT2
835	MSH2	RASA1
836	MSH2	SEC61B
837	MSH2	SFRP4
838	MSH2	TMEFF1
839	MSH2	XPA
840	MSH6	ANP32B
841	MSH6	BAAT
842	MSH6	FBP1
843	MSH6	FBP2
844	MSH6	SMAD1
845	MSH6	STX17
846	MSH6	TBC1D2
847	MSH6	TGFB1
848	MSH6	ZNF510
849	MUTYH	CDC14B
850	MUTYH	DVL1
851	MUTYH	HEMGN
852	MUTYH	IGFBP3
853	MUTYH	MAP2K1
854	MUTYH	PPP3R2
855	MUTYH	SFRP4
856	MUTYH	TMEFF1
857	MUTYH	TMOD1
858	MUTYH	ZNF510
859	NRAS	ANP32B

860	NRAS	BAAT
861	NRAS	IGFBP3
862	NRAS	NANS
863	NRAS	SEC61B
864	NRAS	TGFB1
865	NRAS	TMOD1
866	NRAS	ZNF510
867	ODC1	AKT1
868	ODC1	ALDOB
869	ODC1	ANP32B
870	ODC1	C9orf30
871	ODC1	CDK8
872	ODC1	FRAT2
873	ODC1	GALNT12
874	ODC1	SFRP4
875	PDGFRL	ANP32B
876	PDGFRL	C9orf156
877	PDGFRL	C9orf3
878	PDGFRL	C9orf30
879	PDGFRL	C9orf97
880	PDGFRL	CDC14B
881	PDGFRL	CDK8
882	PDGFRL	CORO2A
883	PDGFRL	CYLC2
884	PDGFRL	DVL1
885	PDGFRL	FBP1
886	PDGFRL	GALNT12
887	PDGFRL	HEMGN
888	PDGFRL	HSD17B3
889	PDGFRL	RASA1

890	PDGFRL	SEC61B
891	PDGFRL	SHC1
892	PDGFRL	STX17
893	PDGFRL	TDRD7
894	PDGFRL	TMEFF1
895	PDGFRL	TMOD1
896	PIK3CA	ANP32B
897	PIK3CA	CDC14B
898	PIK3CA	HSD17B3
899	PIK3CA	KIAA1529
900	PIK3CA	LEF1
901	PIK3CA	MAPK3
902	PIK3CA	SFRP2
903	PIK3CA	SMAD1
904	PIK3CA	STX17
905	PIK3CA	TBC1D2
906	PIK3CA	TMEFF1
907	PIK3CA	XPA
908	PIK3CA	ZNF510
909	PMS2	BAAT
910	PMS2	CYLC2
911	PMS2	FBP1
912	PMS2	HEMGN
913	PMS2	HSD17B3
914	PMS2	IGFBP3
915	PMS2	MAPK3
916	PMS2	NANS
917	PMS2	SEC61B
918	PMS2	TMOD1
919	PTEN	CDK8

Appendix 8. List of interactions associated to CRC detected by Y2H matrix screens

920	PTEN	CYLC2
921	PTEN	FBP1
922	PTEN	FRAT2
923	PTEN	HEMGN
924	PTEN	HSD17B3
925	PTEN	MAPK3
926	PTEN	PPP3R2
927	PTEN	RASA1
928	PTEN	XPA
929	PTPN12	BAAT
930	PTPN12	C9orf156
931	PTPN12	C9orf3
932	PTPN12	C9orf97
933	PTPN12	CDC14B
934	PTPN12	CYLC2
935	PTPN12	FBP1
936	PTPN12	FBP2
937	PTPN12	GALNT12
938	PTPN12	HEMGN
939	PTPN12	PPP2CB
940	PTPN12	RASA1
941	PTPN12	SFRP2
942	PTPN12	TDRD7
943	PTPN12	TMOD1
944	PTPN12	ZNF189
945	PTPRJ	DVL1
946	PTPRJ	HSD17B3
947	PTPRJ	KIAA1529
948	PTPRJ	SEC61B
949	PTPRJ	SFRP2
950	PTPRJ	STX17
951	PTPRJ	ZNF189
952	RAD54B	CORO2A
953	RAD54B	DVL1
954	RAD54B	FANCC
955	RAD54B	FBP2
956	RAD54B	FRAT2
957	RAD54B	GALNT12
958	RAD54B	HEMGN
959	RAD54B	PPP2CB
960	RAD54B	PPP3R2
961	RAD54B	STX17
962	RB1	AKT1
963	RB1	ALDOB
964	RB1	C9orf97
965	RB1	CDK8
966	RB1	CYLC2
967	RB1	FRAT2
968	RB1	HEMGN
969	RB1	HRAS
970	RB1	HSD17B3
971	RB1	PPP3R2
972	RB1	SFRP4
973	RB1	SMAD1
974	RB1	TDRD7
975	RB1	ZNF189
976	RB1	ZNF510
977	SMAD2	FANCC
978	SMAD2	TGFB1
979	SMAD4	ANP32B
980	SMAD4	C9orf156
981	SMAD4	DVL1
982	SMAD4	FBP1
983	SMAD4	GALNT12
984	SMAD4	HSD17B3
985	SMAD4	IGFBP3
986	SMAD4	LEF1
987	SMAD4	MAP2K1
988	SMAD4	SEC61B
989	SMAD4	TDRD7
990	SMAD4	TMEFF1
991	SMAD4	TMOD1
992	SMAD4	XPA
993	SRC	CYLC2
994	SRC	HRAS
995	SRC	SFRP4
996	SRC	SMAD1
997	SRC	ZNF510
998	STK11	BAAT
999	STK11	C9orf97
1000	STK11	CDK8
1001	STK11	FANCC
1002	STK11	GALNT12
1003	STK11	HEMGN
1004	STK11	HSD17B3
1005	STK11	SMAD1
1006	TGFBR2	ANP32B
1007	TGFBR2	BAAT
1008	TGFBR2	C9orf97
1009	TGFBR2	CTSL2
1010	TGFBR2	DVL1
1011	TGFBR2	FBP1
1012	TGFBR2	FBP2
1013	TGFBR2	NANS
1014	TGFBR2	RASA1
1015	TGFBR2	SFRP2
1016	TGFBR2	SFRP4
1017	TGFBR2	TMOD1
1018	TGFBR2	XPA
1019	TLR2	BAAT
1020	TLR2	C9orf3
1021	TLR2	CORO2A
1022	TLR2	CYLC2
1023	TLR2	FBP2
1024	TLR2	HSD17B3
1025	TLR2	LEF1
1026	TLR2	PPP2CB
1027	TLR2	SFRP2
1028	TLR2	SFRP4
1029	TLR2	TBC1D2

Appendix 9. List of interactions associated to CRC detected by yeast two-hybrid library screens

Interacting pairs are reported as gene symbols taken from HGNC (http://www.genenames.org).

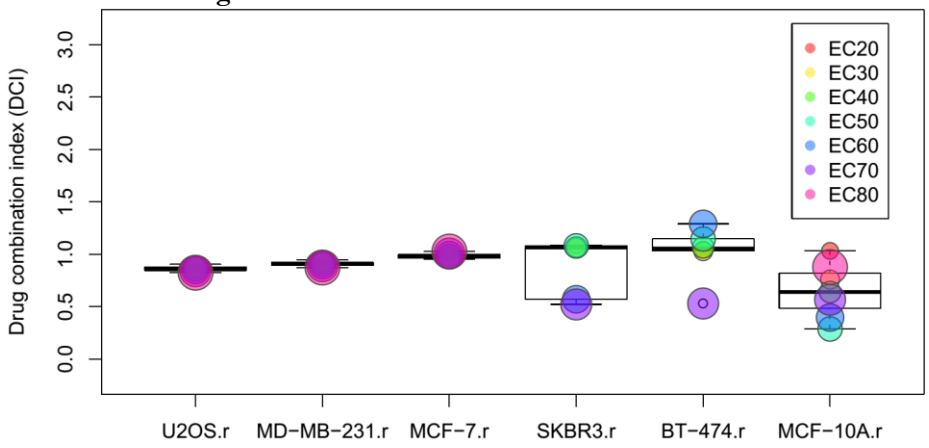
	Bait symbol	Prey symbol	Prey gene name	Blast link (Gene ID)	Prey Uniprot
1	AXIN2	KCNK12	potassium channel, subfamily K, member 12	56660	Q9HB15
2	AXIN2	PRDX1	peroxiredoxin 1	5052	Q06830
3	AXIN2	RANGAP1	Ran GTPase activating protein 1	5905	P46060
4	C9orf30	GNAS	GNAS complex locus	2778	P63092
5	C9orf30	PFN2	profilin 2	5217	P35080
6	C9orf30	TMEM129	transmembrane protein 129	92305	A0AVI4
7	C9orf30	VSTM2A	V-set and transmembrane domain containing 2A	222008	B5MC94
8	DLC1	ACOT7	acyl-CoA thioesterase 7	11332	O00154
9	DLC1	PFKM	phosphofructokinase, muscle	5213	P08237
10	DLC1	PKM2	pyruvate kinase, muscle	5315	P14618
11	DLC1	RANGAP1	Ran GTPase activating protein 1	5905	P46060
12	DLC1	TMEM129	transmembrane protein 129	92305	A0AVI4
13	DLC1	USP4	ubiquitin specific peptidase 4 (proto-oncogene)	7375	Q08AK7/Q13107
14	PDGFRL	ASPH	aspartate beta-hydroxylase	444	Q9H291
15	PDGFRL	MAPK12	mitogen-activated protein kinase 12	6300	P53778
16	PDGFRL	MBP	myelin basic protein	4155	P02686

Appendix 9. List of interactions associated to CRC detected by yeast two-hybrid library screens

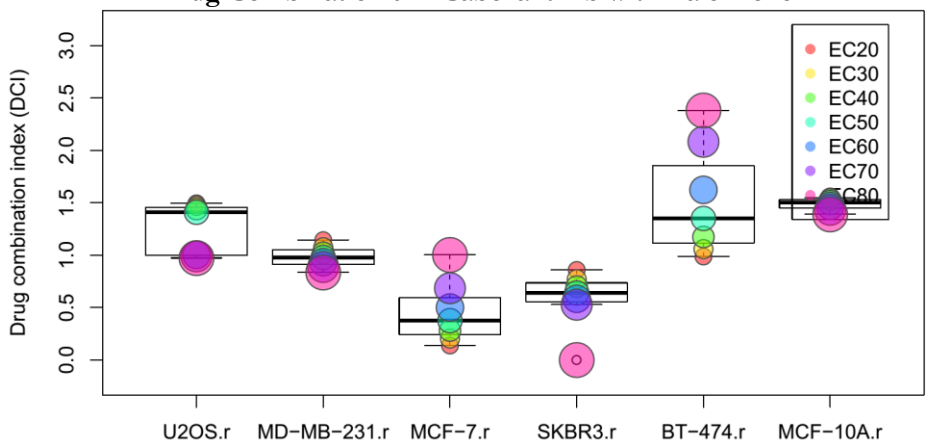
17	PDGFRL	PFKM	phosphofructokinase, muscle	5213	P08237
18	PDGFRL	RANGAP1	Ran GTPase activating protein 1	5905	P46060
19	PDGFRL	SPARCL1	SPARC-like 1 (hevin	8404	Q14515
20	PDGFRL	TMEM129	transmembrane protein 129	92305	A0AVI4
21	SFRP2	DBNDD2	dysbindin (dystrobrevin binding protein 1) domain containing 2	55861	Q9BQY9
22	SFRP2	EEF1G	eukaryotic translation elongation factor 1 gamma	1937	P26641 /Q53YD7
23	SFRP2	PFKM	phosphofructokinase, muscle	5213	P08237
24	SFRP4	AP2M1	adaptor-related protein complex 2, mu 1 subunit	1173	Q96CW1
25	SFRP4	CKMT1B	creatine kinase, mitochondrial 1B	1159	P12532
26	SFRP4	NARF	nuclear prelamin A recognition factor	26502	B3KPX2
27	SFRP4	PFKM	phosphofructokinase, muscle	5213	P08237

Appendix 10. DCI20-80 of drug-drug and target-drug combinations

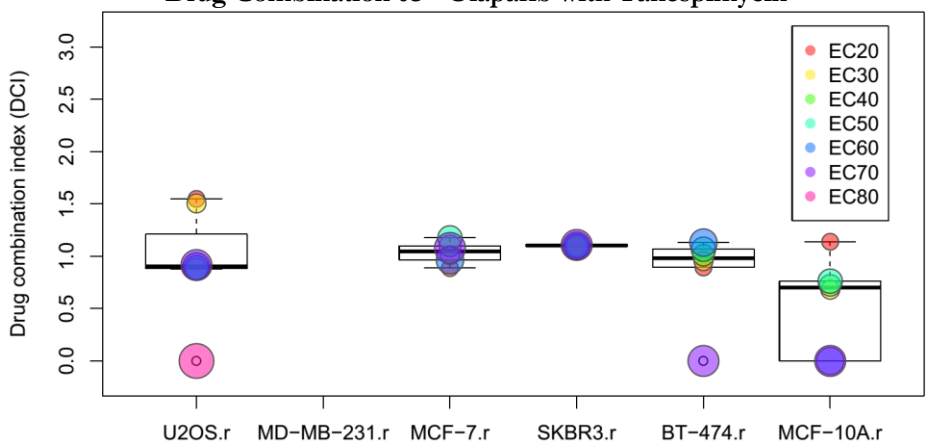
Drug Combination 01 - Cabozantinib with Erlotinib



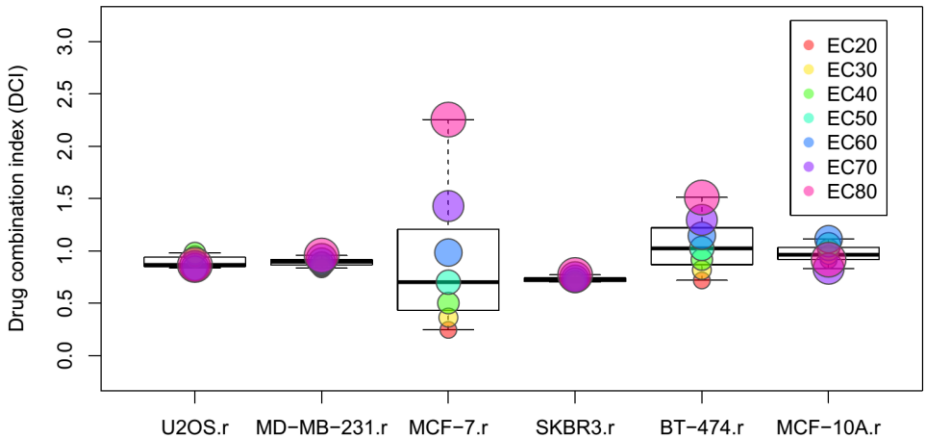
Drug Combination 02 - Cabozantinib with Raloxifene



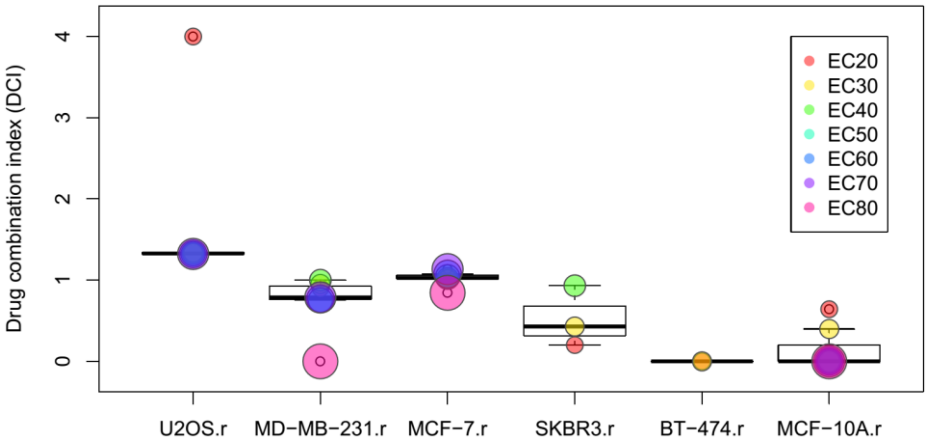
Drug Combination 03 - Olaparib with Tanespimycin



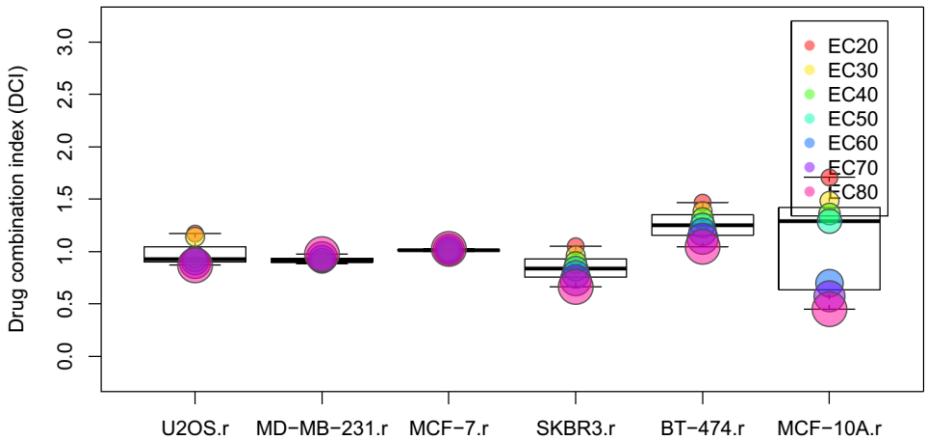
Drug Combination 04 - Olaparib with Dinaciclib



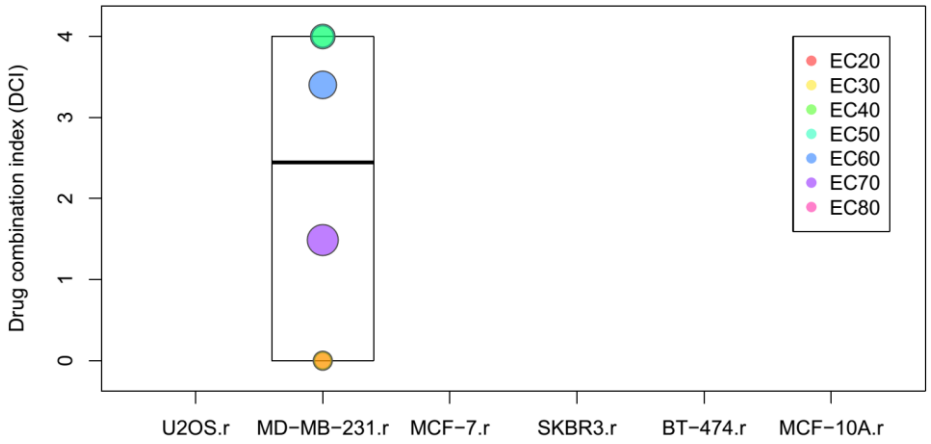
Drug Combination 05 - Olaparib with Palbociclib



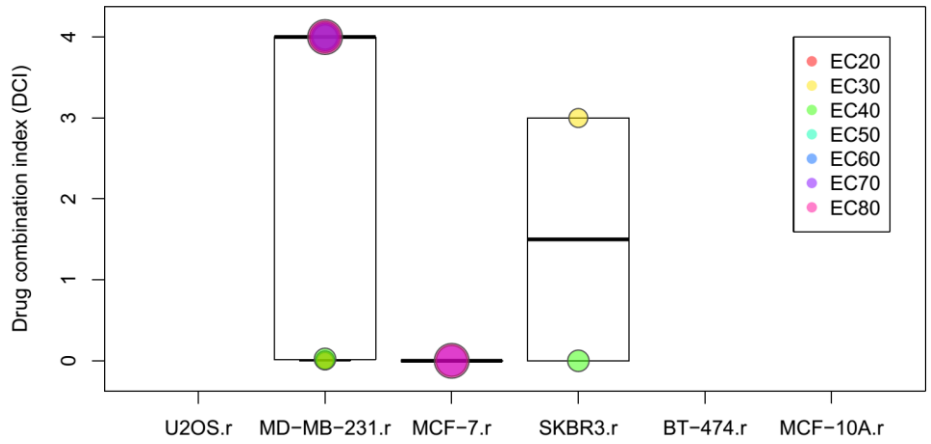
Drug Combination 06 - Cabozantinib with Palbociclib



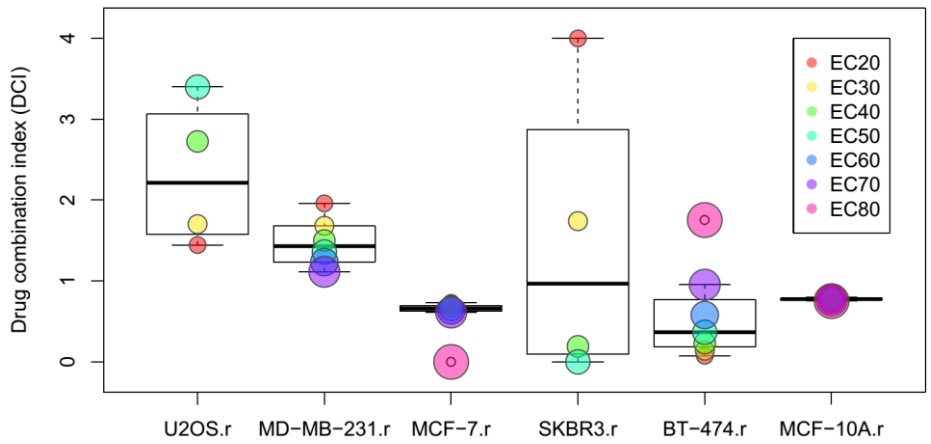
Drug Combination 07 - Paclitaxel with Tanespimycin



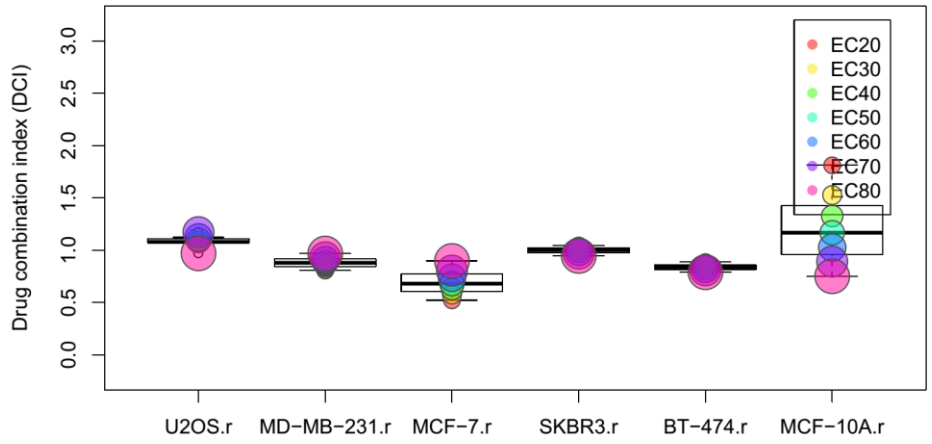
Drug Combination 08 - Paclitaxel with Midostaurin



Drug Combination 09 - Cabozantinib with Trastuzumab

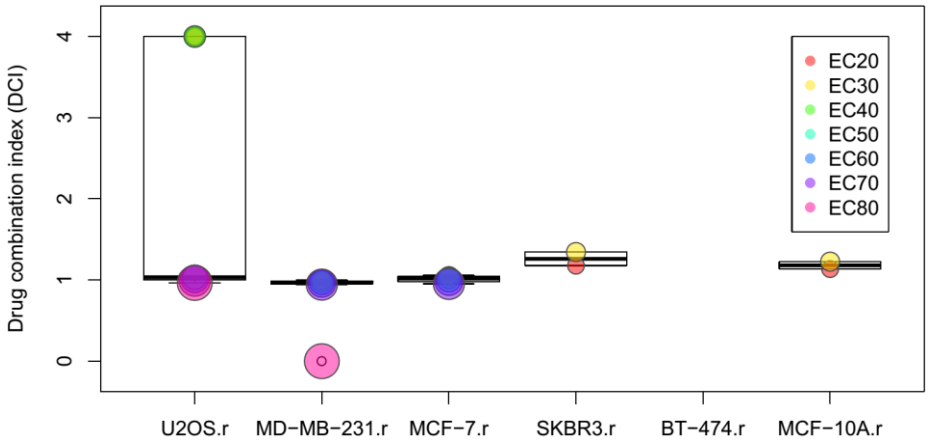


Drug Combination 10 - NVP-AEW541 with Raloxifene

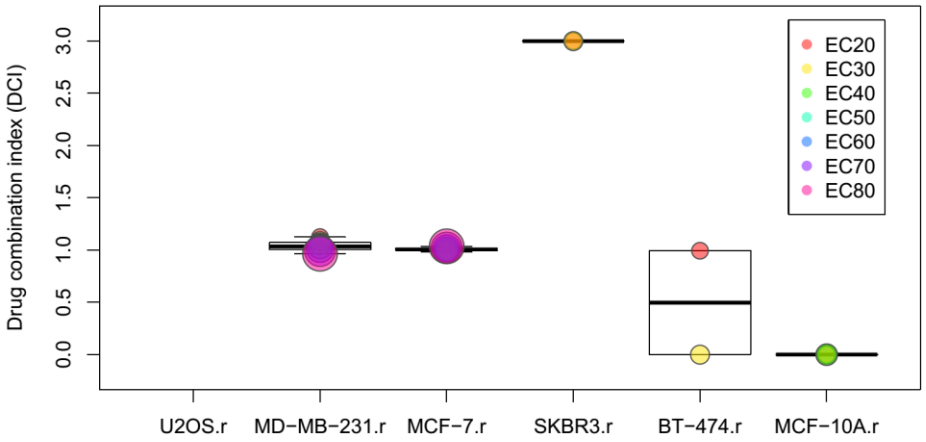


TARGET DRUG COMBINATIONS

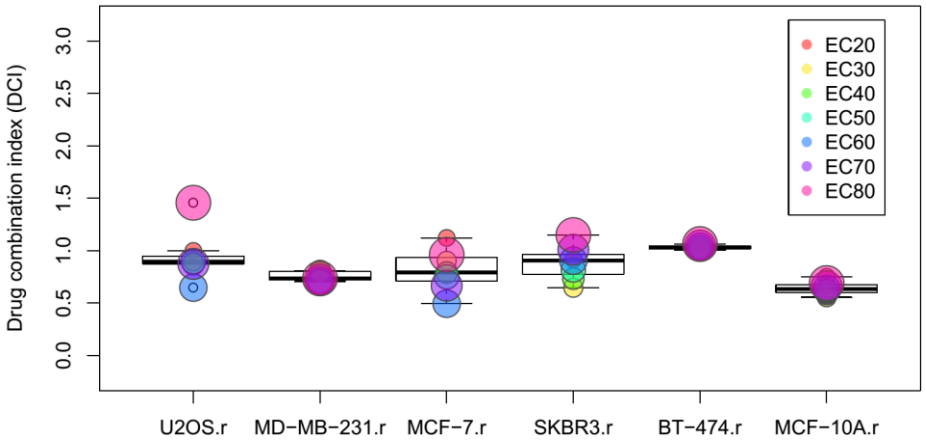
Target Drug Combination 01 - PTPN6 with Olaparib



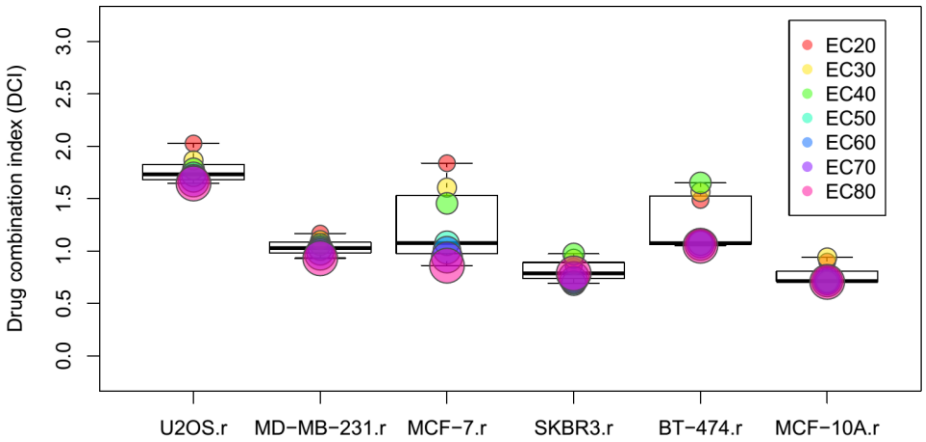
Target Drug Combination 02 - PTPN6 with Palbociclib



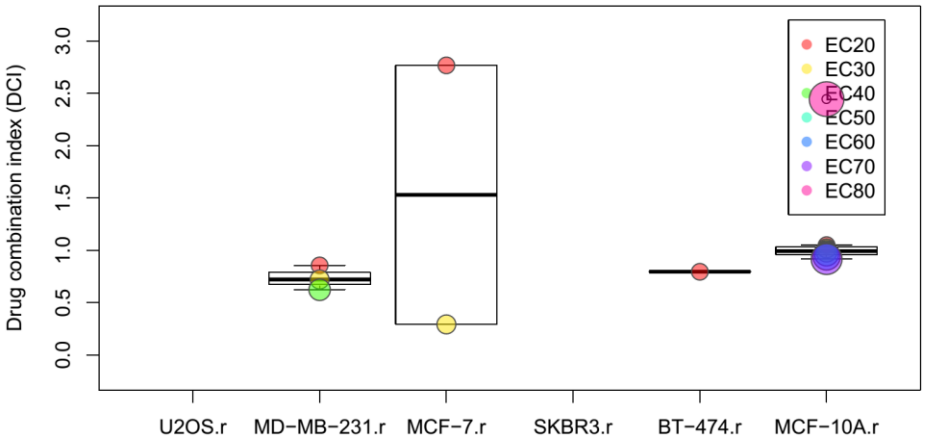
Target Drug Combination 03 - MAP3K7 with Olaparib



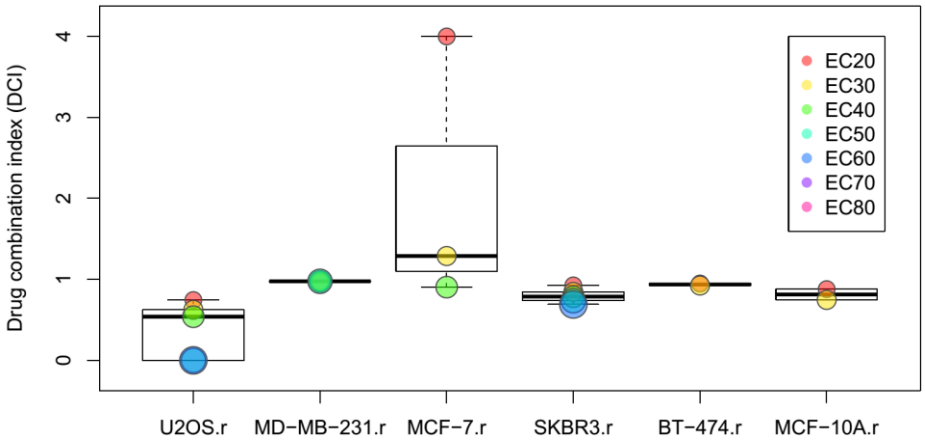
Target Drug Combination 04 - MAP3K7 with Tanespimycin



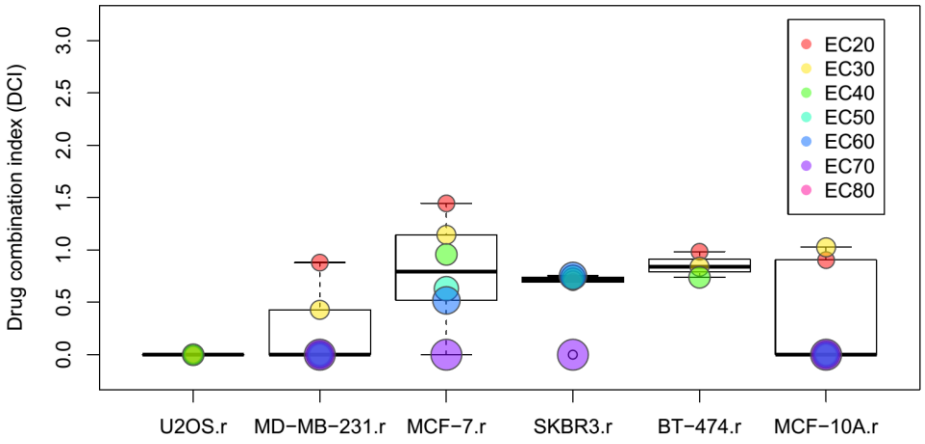
Target Drug Combination 05 - MAP2K2 with Motesanib



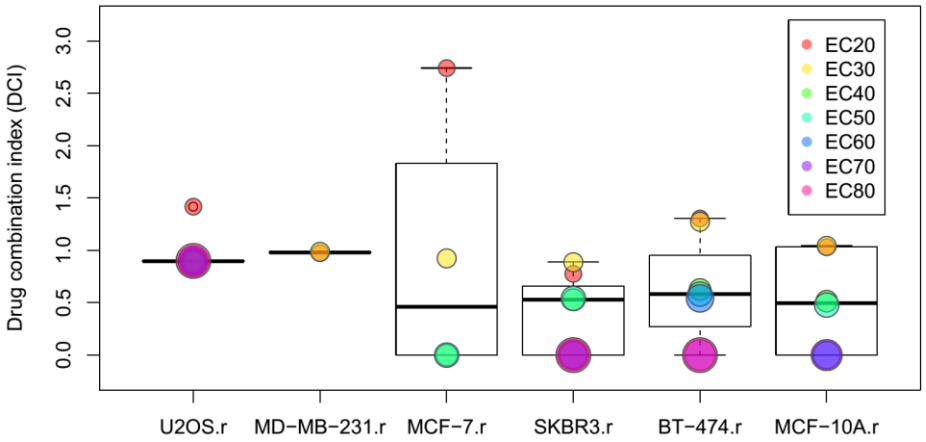
Target Drug Combination 06 - PIK3CB with Motesanib



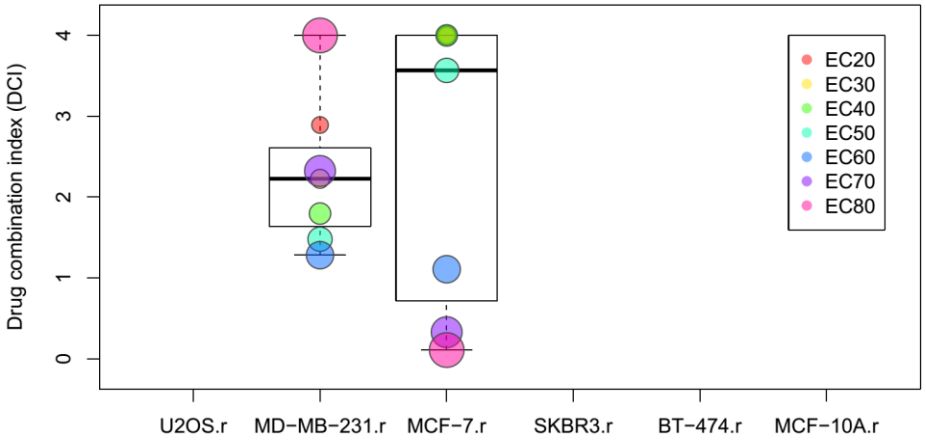
Target Drug Combination 07 - PIK3CB with Palbociclib



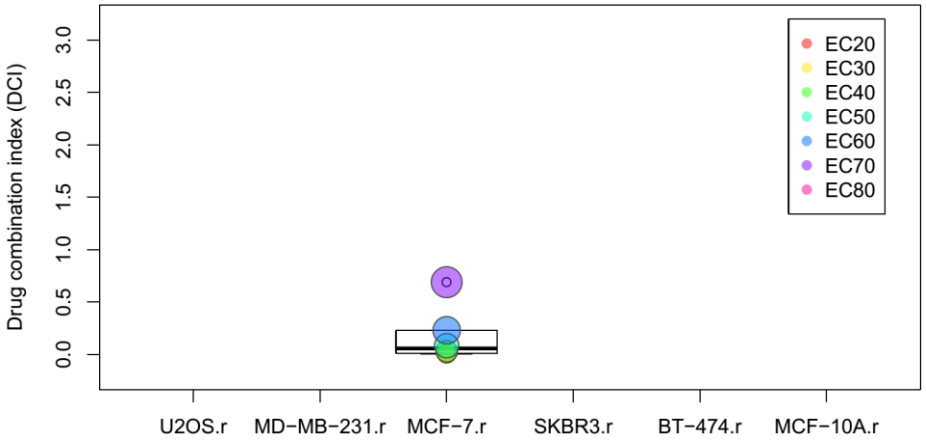
Target Drug Combination 08 - PIK3CB with Tanespimycin



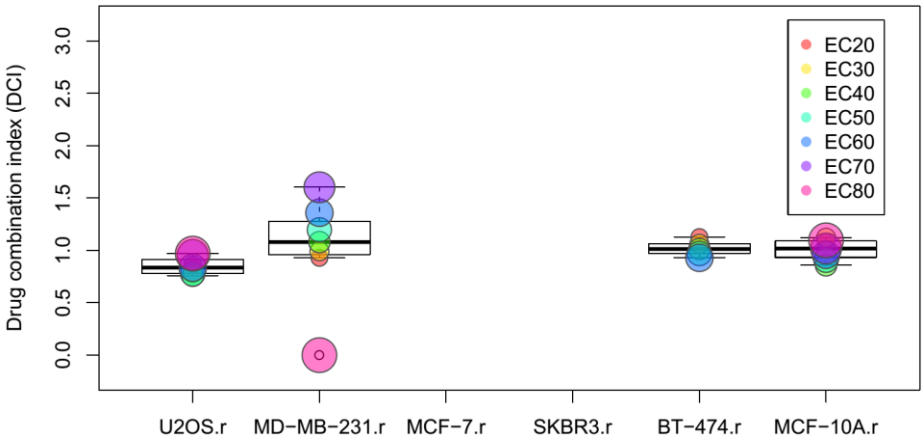
Target Drug Combination 09 - MAPK7 with Paclitaxel



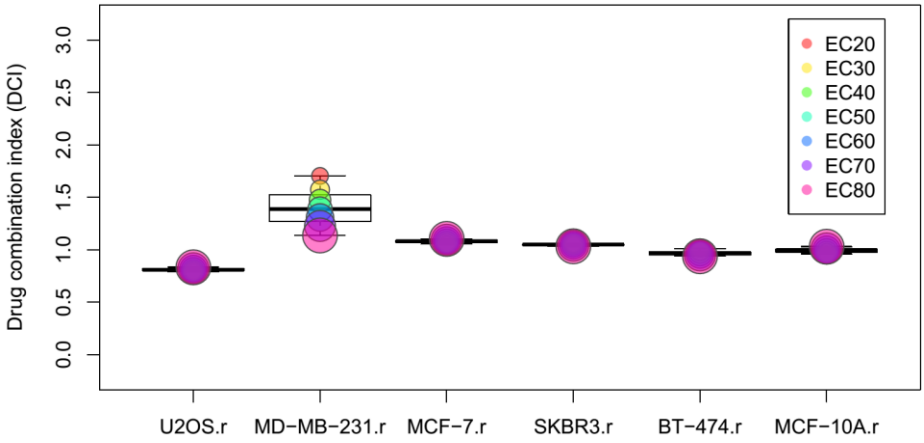
Target Drug Combination 10 - IL1R1 with Paclitaxel



Target Drug Combination 11 - PPP2R5A with Olaparib



Target Drug Combination 12 - PTPN6 with Cediranib



Target Drug Combination 13 - RAP1A with Palbociclib