UNIVERSITÁ DEGLI STUDI DI CATANIA

DIPARTIMENTO DI SCIENZE BIOMEDICHE E BIOTECNOLOGICHE

XXXIV Ciclo di Dottorato

Basic and Applied Biomedical Sciences

Anna Provvidenza Ester Privitera

Driver genes associated to Broad Copy Number Aberrations in solid cancers

TESI DI DOTTORATO

Relatore:

Chiar.mo Prof. **Daniele F. Condorelli**

XXXIV Ciclo 2018/21

A mio Padre

"...l'uomo cavalca il suo alato destriero
in una lunga luminosa scia da favola"

dott. Giuseppe Privitera

INDEX

1. INTRODUCTION

1.1 The role of the aneuploidy in cancer

Aneuploidy is described as an alteration in number of chromosomes (at whole chromosome or chromosome arm level). It is well known its implication in tumorigenesis as a defective segregation mechanism during cell division and, even though it is the oldest cancer alteration found and described by Theodor Boveri (Boveri, 1902) around one century ago, nowadays, it is still unknown if aneuploidy is a direct cause of cancer or malignant transformation. Theodor Boveri between 1904 and 1014 induced multipolar spindles during mitosis (multipolar spindle formation with three or four poles) in sea urchins. The progeny of treated urchins showed aneuploidy. This observation suggested that tumours originated from normal cells with aneuploidy or aberrant chromosomal distribution as effect of aberrant mitosis (Weaver et al., 2006). The **Figure 1** show the aberrant mitosis and their effects on the cell progeny. The Boveri´s theory was the first one about causal connection between aneuploidy and cancer. The scientific data reveals that aneuploidy occur in solid tumours, even thought, less commonly of tumours with near-diploid karyotypes(Weaver et al., 2006).

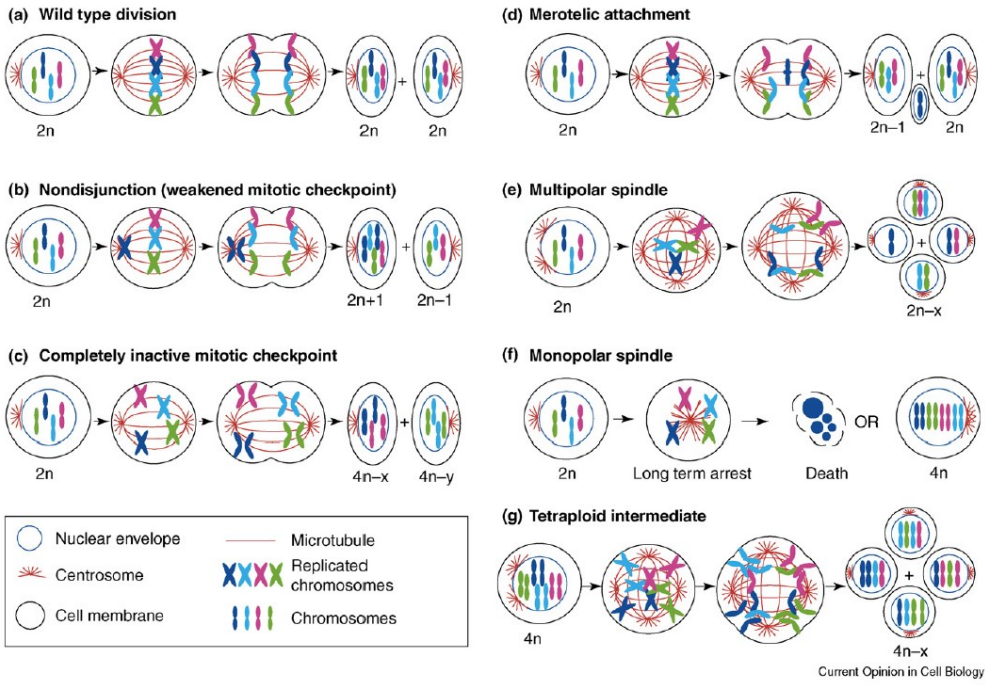


Figure 1. Mechanism generating aneuploidy (Weaver et al., 2006) Copyright © 2006 Elsevier Ltd. All rights reserved.

Aneuploidy is an intrinsic feature of tumour cells for four important reasons: 1) it precedes cell transformation 2) it is observed in pre-cancerous lesions of many different cancer types such as cervix (Duensing & Münger, 2004; Ried et al., 1999), head and neck (Ai et al., 2001), colon (Cardoso et al., 2006; Ried et al., 1999), oesophagus (Doak et al., 2004) and bone marrow (Amiel et al., 2005) 3) it is observed in premalignant lesions in experimental animals such as breast (Medina, 2002) and skin (Dooley et al., 1993) 4) it modifies the transcription equilibrium bringing to both upregulation of those genes promoting cell growing and proliferation, and to downregulation of those genes controlling and reducing the cell growth.

In general, the cells have an innate mechanism to ensure a correct chromosome segregation through the so-called mitotic checkpoint. Mutation and alteration in this mechanism can lead to chromosome instability, nondisjunction errors and near-diploid aneuploidy in which copies of one or a few replicated chromosomes are deposited in the same daughter (Weaver et al., 2006). The most important genes component of the mitotic checkpoint are Mitotic checkpoint serine/threonine-protein kinase (BUB family) and MAD genes in which a mutation, alteration, elimination lead to aneuploidy and chromosome mis-segregation (Weaver et al., 2006).

Any nonconformity from the 2*n* normal karyotype of any region in the genome is a copy number alteration (CNA). Historically, gains or losses of whole chromosomes have been called aneuploidy or whole-chromosome-aneuploidy, while gains or losses involving chromosomal arms or large chromosomal regions have been generally called partial/segmental/structural aneuploidy (Ben-David & Amon, 2020).

Generally, CNAs refer to post-zygote alterations and then in somatic cell and they are termed as somatic copy number alterations (SCNA). While CNAs occurring in the germline, are better termed as copy number variants (CNVs) and they are inheritable. Summarizing, amplifications of chromosomal portions, entire chromosomal arms or entire chromosomes, are generically called gains, while deletions of portions / arms / entire chromosomes are called losses. The cancer chromosomal aberrations (amplifications or deletions) can be subdivided into ***focal*** (small size < 10% of a chromosomal arm) and ***broad*** (whole-chromosome or arm-level aberrations), and it was suggested that focal and broad Copy Number Aberrations (CNAs) can have different consequences for tumour biology (Mermel et al., 2011). In contrast to focal CNAs, few arm-level CNAs show high changes (>3) in copy number (Mermel et al., 2011). The cancer phenotype and cancer progression are consequences of disequilibrium in transcript levels correlated to gene dosage effect (Condorelli et al., 2018). **Broad Copy Number Aberrations** (BCNAs) is another term to

collectively indicate a special type of chromosomal aberrations including classical *numerical aneuploidy* (i.e. an incorrect number of one or more chromosomes like monosomy, trisomy, or tetrasomy), and *structural or segmental aneuploidy* (*gains* or *losses* of chromosomal arm as well as of large chromosomal segments with a size > 50% chromosomal arm) (Mermel et al., 2011; Zack et al., 2013). It is supposed an important role of BCNA in cancer progression although, the majority part of cancer research is focused on focal CNAs (Condorelli et al., 2018). An important concept in tumorigenesis regards abilities and capabilities that cancers acquired and that are shared by all tumours type is about the hallmarks of cancer. The first authors to review these topics were Douglas Hanahan and Robert A Weinberg in their articles "The hallmarks of Cancers" (Hanahan & Weinberg, 2000) and "Hallmarks of Cancer: The Next Generation" (Hallmarks of Cancer: The next Generation, 2011).

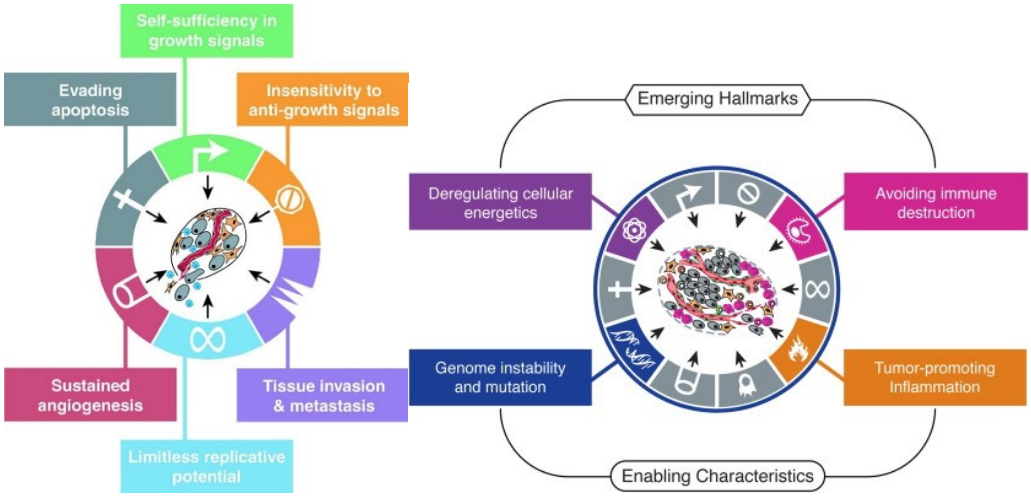


Figure 2 Hallmarks of Cancer. Copyright © 2022 Elsevier Inc. (Hanahan & Weinberg, 2000; Hallmarks of Cancer: The next Generation, 2011)

The **figure 2** shows the hallmarks and the emerging hallmarks and enabling characteristics. Among the hallmarks tumours development we found 1) insensitivity to growth-inhibitory (antigrowth) signals 2) evasion of programmed cell death (apoptosis) 3) self-sufficiency in growth signals 4) sustained angiogenesis 5) limitless replicative potential and 6) tissue invasion and metastasis. Other emerging hallmarks (not yet validated) are: 6) avoiding immune destruction 7) tumor-promoting inflammation 8) deregulating cellular energetics (reprogramming of cellular metabolism) 9) genome instability and mutation.

Usually the normal cells may go through a genetic transformation getting random mutations, which do not provide any contribution to the acquisition of cancer hallmarks. Such mutations are called **passenger mutations**. However, the same process can randomly generate other mutations, called **driver mutations,** that carry cells to a cancer progression with the activation of cancer hallmarks.

The driver mutations are usually the same in different patients, the passengers are all different, so it is believed that those are neutral, and they play no role in cancer.

While the strategy for distinction between cancer driver and passenger mutations is relatively well-established for single-gene or oligo-gene mutational events (such as single nucleotide substitutions, small indels, gene fusions, focal high level amplifications, and focal epigenetic modifications), the task is more difficult for broad genomic aberrations that simultaneously affect several adjacent genes by copy-number dependent mechanisms (Somatic Broad Copy Number Aberrations: BCNAs). In the context of the *classical two-hit mechanism* for inactivation of **tumor suppressor genes** (TSG), providing that the first allele has been inactivated by a single-gene mutational events (inactivating point mutation or small focal deletion), the remaining functioning allele can be removed by a broad genomic loss (for instance a chromosome monosomy). Along this line of reasoning, the accompanying hemizygous loss of several genes contained in the same genomic segment can be considered as passenger mutations, i.e. neutral for the evolution towards a cancer phenotype. In contrast with this simple view, it is conceivable that the simultaneous loss or gain of several genes contained in broad deleted or amplified chromosomal regions can also exert positive (driver) or negative (deleterious) effects for cancer phenotype selection (Roy et al., 2016; Solimini et al., 2012). The main role of BCNAs, and genes related to them, is better clarified from three relevant inactivation mechanisms of TSGs: haploinsufficiency (monoallelic inactivation) and biallelic inactivation (Davoli et al., 2013). In general TSGs need a second hit (biallelic inactivation) to produce cancer phenotype. Even so, a monoallelic inactivation of TSGs could induce a phenotypic modification independently by a second hit by haploinsufficiency (Condorelli et al., 2018). If more than one TSG show haploinsufficiency there is a cumulative haploinsufficiency effect which is enough to develop cancer. It has been suggested that a large majority of TSGs involved in sporadic cancer are haploinsufficient (Davoli et al., 2013). A similar approach can be applied to the so called “triplosensitivity”, indicating the increment of gene copies (oncogenes) due to a contextual increment of chromosomal portion or whole chromosome (Davoli et al., 2013).

1.2 Numerical and structural chromosomal aberrations

Historically, from a cytogenetics point of view, the chromosomal aberrations are classified in two large groups: numerical and structural aberrations. A Numerical chromosomal aberration refers to a change in the normal number of chromosomes as a missing (monosomy) or gaining (trisomy, tetrasomy) of portions (p and q chromosomal arm) or whole chromosomes. A structural chromosomal aberration refers to a change in the structure or shape of chromosomes. The most known numerical chromosome aberrations include Monosomy X (Turner´s syndrome), Monosomy 5p (Cri du chat syndrome), 1p36 Monosomy syndrome, Trisomy 18 (Edward's syndrome) and Trisomy 21 (Down syndrome). The structural chromosomal aberrations are several types including: 1) **Deletions**: loss of genetic material involving a single missing base pair or piece of the chromosome 2) **Inversion**: Segment of chromosomes come up against a double strand breaking, followed by a turning around and a reattaching to itself. On the base of the breaking point around the centromere, the inversions are classified in pericentric and paracentric. As rule parents can pass on inversions to their offspring, but they don't always result in syndrome or birth defects. 3) **Translocation**: exchange of chromosomal portions or entire arms belonging, mainly, to different chromosomes (non-homologous exchanges). Translocations can be reciprocal with portions of chromosomal arms exchanged and resulting in two different derived chromosomes, or non-reciprocal where the transfer of chromosomal material between the arms is unbalanced. If no genetic material is acquired or lost, it is called a balanced translocation, in contrast to unbalanced translocations in which the gain and / or loss of genetic material is observed. 4) **Rings chromosome**: A chromosome bends and fuse together, forming a circular ring. Sometimes rings chromosome shows a loss of genetic material, in other instances they do not.
Chromosomal translocations are very common phenomena in human cancer, particularly in hematopoietic and lymphoid tumours (Mitelman et al., 2007). Most common structural chromosomal alterations found in human carcinomas and hematopoietic tumours are inversions (pericentromeric or centromeric ruptures), translocations, deletions and gains (Ferguson et al., 2001). In oncogenic chromosomal translocations, rearrangements can alter the original positions of proto-oncogenes, i.e. those genes that contribute to stimulating cell proliferation and division, through two main modes: 1) Generating oncogenic fusion proteins 2) Transferring proto-oncogenes in proximity of cis-regulatory elements (promoters, enhancers) (Zheng, 2013).

Usually, when the translocation generates oncogenic fusion proteins it is possible to justify the pathogenesis of cancer, but in many other cases the translocation does not produce any aberrant products and it is difficult to understand the exact biological role in the pathogenesis of cancer. Since no alterations of single genes have been revealed at or near the breakpoint junctions of several unbalanced chromosomal aberrations, there is a wide consensus that gene-dosage transcriptional effects are playing a role in tumorigenesis and cancer progression. As it will be discussed later this is the case of the derivative (1,16) in breast cancer.

1.3 The cancer genome atlas (TCGA)

Nowadays many studies are trying to clarify the role of aneuploidy in tumorigenesis and to investigate the possibility to develop tumour-specific aneuploidy models. Studies reveal that ~90% of solid tumours show aneuploidy with a different percentage in according to the tumour type (Taylor et al., 2018). A milestone in the understanding cancer aneuploidy and chromosome arm-level aneuploidy is the method proposed by Alison M. Taylor et al. (Taylor et al., 2018) by performing a pan-cancer computational analyses on 10.522 tumors of 33 cancer types deposited in the Cancer Genome Atlas (TCGA) database. The 10.522 samples were analyzed for somatic DNA copy number by Affymetrix SNP 6.0 arrays, for expression values by Illumina mRNA sequencing platform and for mutations by whole-exome DNA sequencing platform. Thanks to a bioinformatics approach to manipulate chromosomes, Taylor group calculated the ploidy and absolute total copy number of each chromosome (p and q arm) and each segment in the genome. In general, they considered deleted, neutral or amplified a segment with a copy number smaller, equal or greater than sample's ploidy respectively. While, for the alterations (amplifications/deletions), segments were considered altered if deletion or amplification was more than 20% of the chromosomic length. Successively, they clustered tumours in aneuploid and non-aneuploid and for each chromosome they have assigned a value to calculate the number of altered arms: -1 if lost, 0 if non-aneuploid, +1 if gained, and "NA" if other. The values so assigned were then summed and for each tumor were assigned an score, the so called aneuploidy score which ranged from 0 to 39 (see **Figure 3B**) (the max value corresponds to all arms altered both long and short arms for each non-acrocentric chromosome, and only long arms for chromosomes 13, 14, 15, 21, and 22) (Taylor et al., 2018). The **Figure 3** is a summary of the bioinformatics analysis of the specific aneuploidy pattern of cancer types (**Figure 3A**) by using the Pearson´s method in a hierarchical clustering. For each arm (p e q) of each chromosome is visualized the

corresponding mean arm alteration value (blue = -1/loss, white = 0/neutral and, red = +1/gain). In **Figure 3C** is a graphic representation of the sum of Aneuploidy score with the blue colour are identified the tumours not doubled, with green colour the genome with one-fold genome doubling and with red the genome more than one fold genome doublings). The 33 cancers type (**Figure 3C**) are ordered from the one with less aneuploidy (THCA) to the one with the most aneuploidy (TGCT).

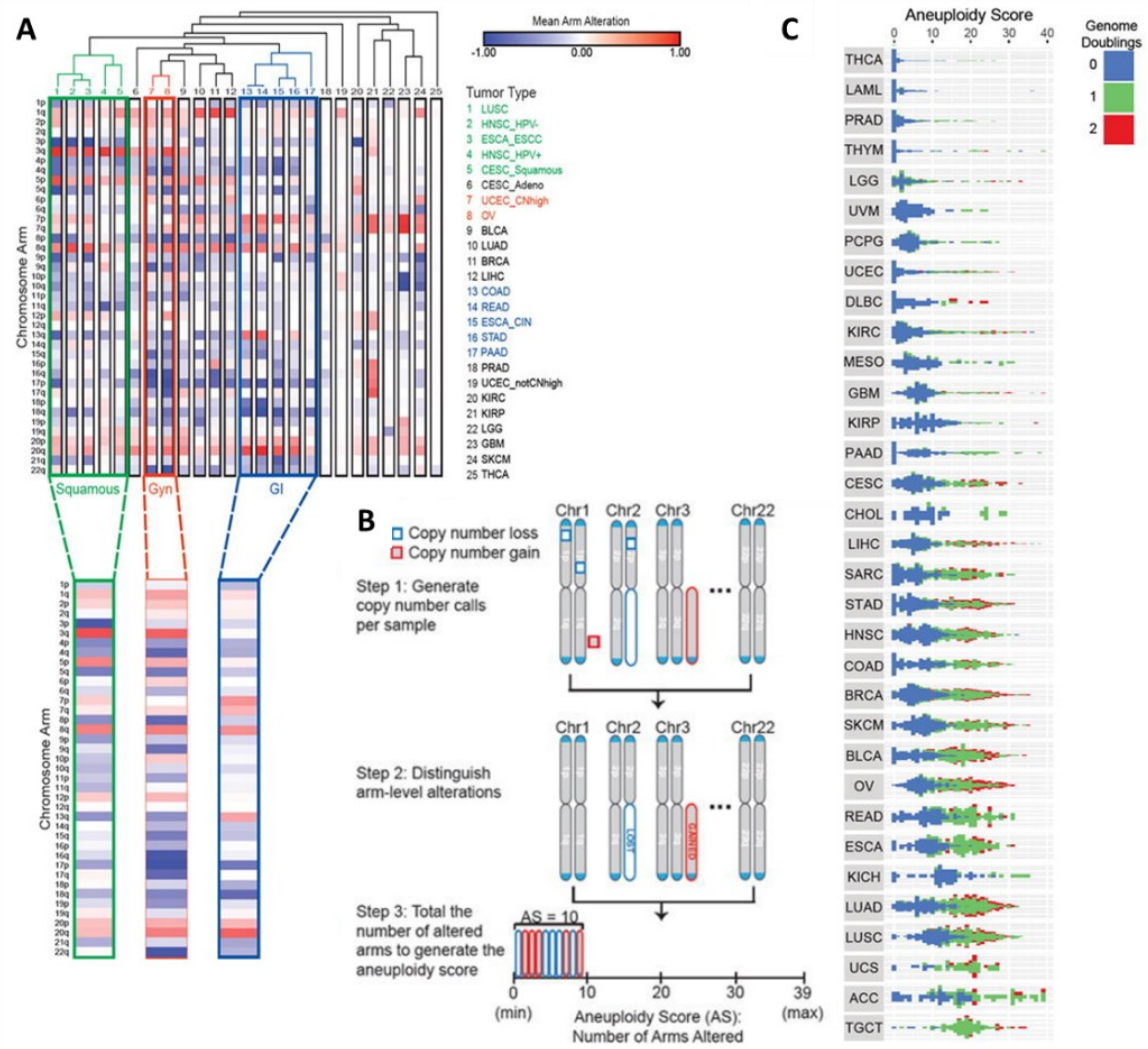


Figure 3: A) Hierarchical clustering showing the mean arm alteration in each chromosome (p and q arm) and for each cancer type B) the schema to attribute the value -1, 0, +1 to calculate the aneuploidy score. C) Graphics representation of the tumours ranked from the one with less aneuploidy (THCA) to the one with the most aneuploidy (TGCT) (Taylor et al., 2018).

The **figure 3C** show how aneuploidy rate varies across cancer types. For example 26% of thyroid carcinomas show arm-level alteration, glioblastomas (99%), uterine carcinosarcomas (96%) and testicular germ cell tumours (99%) (Taylor et al., 2018). Undoubtedly, Taylor´s results convince on the role of aneuploidy in cancer and on the existent of a **cancer-type-specific chromosome arm-level aneuploidies patterns**. Proof of this, Colorectal (COAD-

READ), non-squamous oesophageal (ESCC), stomach (STAD), and pancreatic (PAAD) show gains of chr8q, chr13q, and chr20 (Taylor et al., 2018). In Ovarian cancer (OV), lung adenocarcinoma (LUAD), breast cancers (BRCA), and liver hepatocellular carcinoma (LIHC) show gain of 1q with high sCNAs (Taylor et al., 2018). Low grade glioma (LGG), glioblastoma (GBM), uveal melanoma(UVM), renal clear cell carcinoma (KIRC), and renal papillary cell carcinoma (KIRP) show gain of chr7 with few sCNAs (Taylor et al., 2018). In lung squamous cell carcinoma (LUSC) co-occur loss of chr3p (80% of tumours) and gain of chr3q (60% of tumours)(Taylor et al., 2018).

The aneuploidy produce consequences on the driver genes such as the cancer hallmark genes, instead, when 3p which is deleted across TCGA samples, the cancer hallmark sets of cell cycle (E2F targets/G2M Checkpoint), epithelial mesenchymal transition, interferon gamma response, and TNFa signalling are upregulated (Taylor et al., 2018) . The association between mutational state and aneuploidy was as well exploited. In general, cancers with low aneuploidy scores have high mutational rate(Taylor et al., 2018). In COAD and uterine corpus endometrial carcinoma (UCEC) low aneuploidy cancers with high mutation rate is associated with POLE mutation and microsatellite instability(Taylor et al., 2018).

Even if the Taylor´s study has clarified many aspects of cancer biology, many questions remains unsolved. For example, what is the selective pressure which select the cancer with aneuploidy? Which is the role of the stroma interaction in aneuploidy? These and other unsolved questions need to be exploited through single cells analysis, cellular modelling, as well as a massive and productive computational approach.

1.4 Experimental design to investigate aneuploidies

An experimental design to investigate the broad copy number alterations (BCNAs) (arm-level or whole chromosome aneuploidy) and their effects on transcriptomic profile in cancer is proposed in Condorelli et al. 2018 (**Figure 4**). The model aim is based on 1) identification of BCNAs in a specific type of cancer 2) Grouping the tumour samples with a specific chromosome aberrations (gain or loss of a chromosomal- arm or a whole chromosome) to determinate the so called "*selected CRC group"* 3) Grouping the tumour samples not bearing that specific chromosome aberration to determinate the so called "*Control groups"* 4) Grouping the normal colon mucosae samples 5) Performing computational analysis to identify significant differential expressed genes (DEGs) between selected CRC group and corresponding control CRC group (this comparison determines a group of genes called OverT or UnderT) or between selected CRC group and normal colonic mucosae group (to determine UpT and DownT genes). The consecutive steps regard 6) the evaluation of transcript levels of a given transcript class by the parameter "Chromosomal Distribution Index" (CDI) and 7) Biological function analysis(Condorelli et al., 2018).

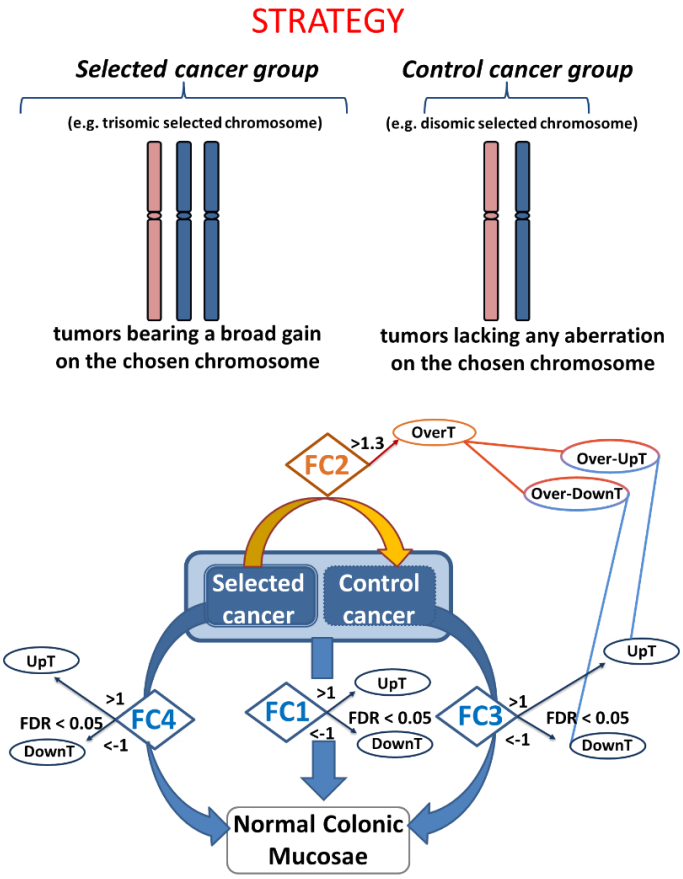


Figure 4: Model to investigate aneuploidies in cancer (Condorelli et al., 2018).

The model discussed was developed to investigate BCNAs in Colon cancer (COAD), nevertheless it is applicable to any cancer type. In this PhD thesis the model was improved by using a normalized chromosomal distribution index (NCDI) instead of an only Chromosomal distribution index (CDI) which takes into account the number of transcripts enriched in a specific "transcript class" but normalized as percentage of transcripts respect to the total number of transcript in one chromosome as better described in material and method section. The computational analysis was much improved developing a bioinformatics pipeline applicable to each cancer type. The range of data has been extended by considering RNA-seq data, SNP-arrays data and mutational data such as whole-exome sequencing.

1.5 Aneuploidies in Breast Cancer

Breast cancer displays recurrent genome copy number aberrations as demonstrated in combined studies of genome copy number and gene expression profiling. In particular recurrent chromosome aberrations regards the gain of 1q, 8q e 16p and the loss of 16q e 17p (Chin et al., 2006; Ried et al., 1999). As reported in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (https: //mitelmandatabase.isb-cgc.org/, accessed on 1 August 2020) as well as in multiplicity studies (Dutrillaux et al., 1990; Farabegoli et al., 2004; Kokalj-Vokac et al., 1993; Muthuswami et al., 2013; N. Pandis et al., 1994; Nikos Pandis et al., 1992, 1995; Russnes et al., 2010; Rye et al., 2015; Teixeira et al., 2002; Tsarouha et al., 1999), in breast cancer, the most common cytogenetics abnormalities refer to 1) the derivative chromosome der(1;16) (q10;p10) 2) the isochromosome 1q (i(1q)) and the deletion of del(16q). The der(1;16) (q10;p10) is considered to be the consequence of a centromere-close translocation t(1;16) and it consists in a gain of long arm of chromosome 1 and a loss of long arm of chromosome 16(**Figure 5-6**). The i(1q) consists in a loss of the entire short arm (p) of chromosome 1 with subsequent duplication of the entire long arm (q) of chromosome 1 (**Figure 5-6**). The del(16q) consists in a deletion of 6q-long arm 1 (**Figure 5-6**).

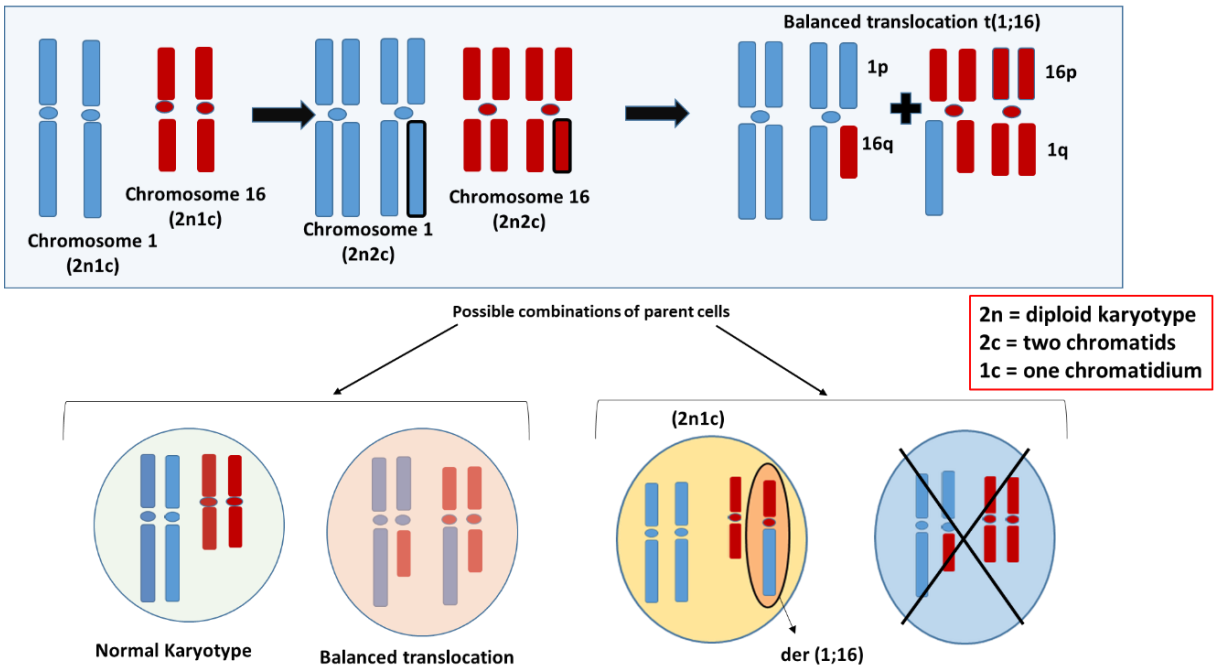


Figure 5: A model for generation of der(1;16) (q10,p10).

However, their exact role in determining the cancer phenotype is still largely unknown. There is a wide consensus that gene-dosage transcriptional effects are playing a role in tumorigenesis of breast cancer by gaining extra copies of chromosome 1.

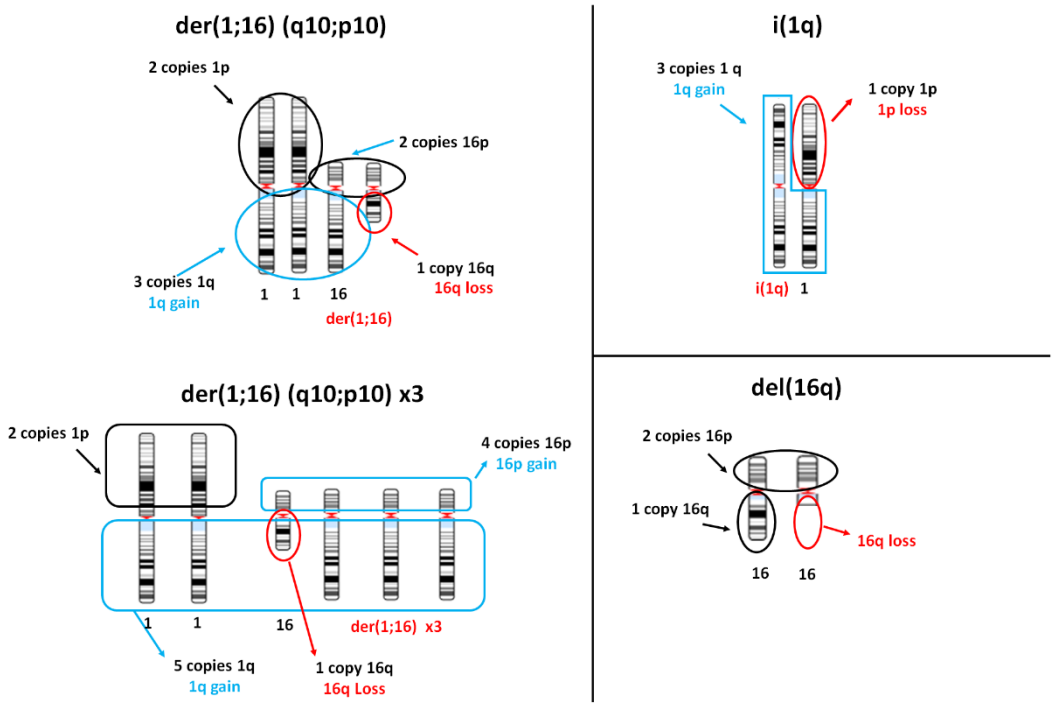


Figure 6: Schematic drawings of main cytogenetic aberrations in breast cancer and corresponding copy number changes. Left panel top: single copy of der(1;16) (q10;p10). Left panel bottom: 3 copies of der(1;16). Right panel top: isochromosome 1q (i(1q)), a chromosome formed by two long arms of chr1. Right panel bottom: deletion of long arm of chr16.

Though the frequent association of 1q-gain and 16q-loss suggests a cooperation between those aberrations, it has been also reported that breast cancers can bear only 1q-gains or only 16-q losses (Curtis et al., 2012; Farabegoli et al., 2004). Through a comparative genomic hybridization (CGH) studies, Farabegoli et al, have individuated 1q23-q32 and the loss of 16q12-qter as chromosomal region involved in the der(1,16) in breast cancer as well as minor amplification events in 7p21-22, 20p11.2 and 20q12-13.1 chromosomal regions (Farabegoli et al., 2004). The breast cancer is classifiable in two large intrinsic subtype the invasive ductal (IDC) and the invasive lobular carcinoma (ILC). The 1q gain is typically present at every stage of carcinoma, while, 16q loss aberrations is detected in differentiated carcinomas or invasive lobular carcinoma (Farabegoli et al., 2004). Considering the molecular classification, breast cancer is subdivided in 5 intrinsic molecular subtypes: Luminal A, Luminal B, Her2, Basal-like and Normal-like, according to the PAM50 intrinsic subtypes classification (Picornell et al., 2019) (Gnant et al., 2014) (Dowsett et al., 2013). The frequency of this 1q gain/16q loss is higher in Luminal A related to HER2-enriched or basal-like. (Rye et al., 2015).

1.6 Breast cancer classification

Breast cancer is a biologically and phenotypically heterogeneous disease with various morphological and molecular patterns which bring to molecular and histological classifications (Nascimento & Otoni, 2020).

1.6.1 Molecular classification

According to the PAM50 (Prediction Analysis of Microarray 50) intrinsic subtypes classification (Picornell et al., 2019) (Gnant et al., 2014) (Dowsett et al., 2013) is based on a gene expression study screening 9000 genes. A list of 550 genes were identified as crucial to discriminate the phenotype of individual tumours (Perou et al., 2000) (Sørlie et al., 2001). In particular, the expression patterns of these genes allow to divide breast tumours in 5 intrinsic molecular subtypes which are Luminal A, Luminal B, Her2, Basal-like and Normal-like.

Tumors can be classified according to ER status in ER+ and ER-. A further classification is given by the expression of the HER2/erbB2/neu receptor protein, i.e. a tyrosine kinase receptor involved in the regulation of cell growth, expressed by about 10-15% of tumors (Russnes et al., 2017). Immunohistochemistry (IHC) techniques also allow to evaluate the expression of estrogen receptors (ER) and progesterone (PgR) and it has been found that about 75% of breast cancers express these receptors (Russnes et al., 2017)

The subtypes Luminal A and Luminal B are dominated by ER positive (ER+) and by genes expressed in luminal breast epithelial cells , cells surrounding the duct lumen (Mueller et al., 2018). The subgroup Luminal B has generally a high expression of gene involved in cell proliferation and mitosis (Russnes et al., 2017). The group called Basal-like, are characterized by genes typically expressed in myoepithelial or basal epithelial cells. Basal-like subtype are ER negative, PgR negative and Her2 negative and known as triple-negative breast carcinoma (TBNC) (Russnes et al., 2017).

From a clinical outcome, tumours ER+ and PR+ have a favourable prognosis, giving the likelihood of a better response to endocrine therapy. Meanwhile tumours with the expression of HER2 are more aggressive, have poor prognosis and low response to chemotherapy, because this receptor is related to growth factor inducing cellular proliferation (Moasser, 2007). On the other hand, the basal-like tumours, specifically triple negative and ER-, have the worst short-term prognosis, even though is less frequent than the other two previously described.

1.6.2 Histological Classification

Tumours used to be generally classified by histopathologic analysis on the base of tubule formation, mitotic counting and the degree of nuclear pleomorphism (Russnes et al., 2017). The most frequent is the invasive ductal carcinoma (IDC) (not otherwise specified), due to lack of sufficient characteristics. The second most frequent is the invasive lobular carcinoma (ILC). Other minor types of breast cancer are apocrine, cribriform and mucinous carcinoma so classified according to the growth patterns.(Weigelt et al., 2008). In **figure 7** it is reported a classification of the main breast cancer histotypes.

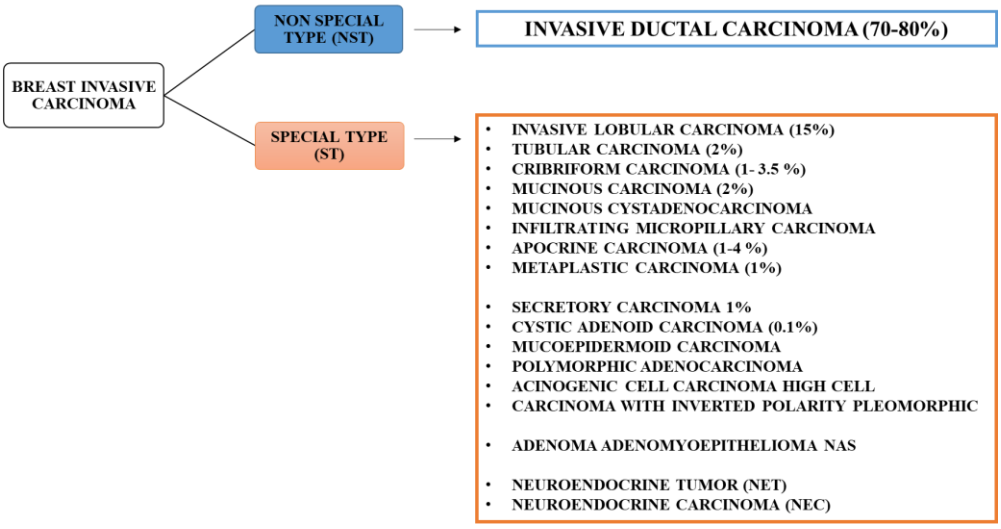


Figure 7: Schema about the main histological subtype.

The histological special types (ST) show a more homogeneous phenotype than no-special type (NST). Many of these tumours are treated with neoadjuvant and chemotherapy but for the secretory and apocrine carcinoma, which represents only 1% and 1-4 % of breast invasive carcinoma special type apocrine carcinoma, there is any efficacy therapy. The secretory carcinoma was originally observed in children and has been referred to as 'juvenile carcinoma' but it can occur at any age. The clinical course is usually excellent, even in case of nodal involvement or metastasis.

The secretory carcinoma shows a tubular, solid and/or microcystic architecture and abundant colloid-like extracellular secretion. The secretory carcinomas show particular gene feature, the ETV6-NTRK3 fusion gene (Pareja et al., 2021). This chimeric gene derived from the t(12;15)(p13;q25) translocation that involves the dimerization domain of the transcriptional regulator (ETV6)(located on chromosome 12) and the membrane receptor tyrosine kinase (NTRK3) (also known as TRKC and located on chromosome 15) (Knezevich et al., 1998) producing a potent tyrosine kinase with high transforming effect (Q. Liu et al., 2000; Wai et

al., 2000) (Wai et al. 2000, Liu et al. 2000). ETV6-NTRK3 expression actives two effector pathways of wild-type NTRK3 which have an impact for cell proliferation and survival in breast cancer: 1) the Ras-MAP kinase (MAPK) mitogenic pathway and 2) the phosphatidyl inositol-3-kinase (PI3K)-AKT pathway mediating cell survival (Tognon et al., 2001; Wai et al., 2000). In particular, ETV6-NTRK3 leads to a constitutive activation of membrane receptor of tyrosine kinases. This observation focus the attention on the NTRK signaling, also known as Trk signaling, to be used as targets for new cancer therapies (Euhus et al., 2002) as well as to hypothesize a genetic target..

Instead, a screening of chemical libraries to find inhibitor of trk signaling, have identified the drug known as Entrectinib, as inhibitor of TRK. (Doebele et al., 2020).

Undoubtedly these results, open the perspective to new cancer therapies as well as to new biomarker discovery for those minor breast cancer types (i.e. secretory carcinoma) for which there are no effective chemotherapy and reliable genomic diagnosis.

1.7 Aneuploidies in Colon Carcinoma

Colon cancer is one of the most common human malignancies and is a leading cause of death worldwide. Incidence of colorectal cancer ranks third, behind prostate cancer and lung cancer, in males, and behind breast cancer and lung cancer in females (Ferlay et al., 2015). The vast majority (over 95%) of the cancers that origin in the colorectal region are classified as adenocarcinomas (Hamilton et al., 2010). At least two forms of genomic instability have been described in colorectal adenocarcinomas: *chromosomal instability* (CIN) and *microsatellite instability* (MSI). CIN is characterized by numerical and structural chromosomal aberrations and is present in the large majority of tumors (about 85%). MSI is identified by somatic changes in the number of repeating units of microsatellite repeats due to defects in the DNA mismatch repair genes. Although MSI and CIN have been considered mutually exclusive, it is now clear that MSI tumors also show-varying degrees of CIN and that a percentage of the tumors can be considered positive for both form of instability (Grady, 2004). In an early study involving 172 colon-rectal samples (40 samples early-stage adenomas, 40 adenomas and 92 carcinoma), allelic deletions of chromosomes 5, 17, and 18 are found. Loss of chromosome 5 was found in 26-35% of adenomas and carcinoma. Loss of a portion of chromosome 18 was deleted in 78 percentage of carcinoma, in 47 percentage of late-stage carcinomas and 11-13 % in carcinoma, while loss of chromosome 17p in 75% of carcinomas (Vogelstein et al., 1988). In colon cancer biology the increase of chromosomal number of copies (chromosomal gain) is a critical topic of considerable interest. Chromosome 20 amplification are common chromosome aberration, as demonstrated by sequencing a cohort of 401 carcinomas and founding a 20q gain in 37% of samples and a 20q amplification in only 7% of samples (Ptashkin et al., 2017). Hotspot mutations in *KRAS* and *BRAF* were well documented in literature and they are found in 10% of colorectal carcinoma bearing 20q amplification and in 45% of colorectal carcinoma bearing 20q gain, and 72% of colorectal carcinoma no bearing aneuploidy of 20q (Ptashkin et al., 2017). It has been observed a causal correlation between increase of chromosomal number of copies and high gene expression because these genes are more dosage sensitive than low-expressed genes (Fehrmann et al., 2015). Taylor et al (Taylor et al., 2018) have studied a negative correlation between hotspot mutations and colorectal carcinoma (COAD and READ), founding a high level of MSI or *POLE* mutation in samples with low aneuploidy rate, demonstrating how aneuploidy is a sufficient aberration for tumorigenesis. A pattern of aneuploidy was proposed by performing a hierarchical clustering with Pearson's correlation. The results showed a co-occurring alteration of chr7p, chr7q,

chr8q, chr13q, and chr20q (Taylor et al., 2018) . In Condorelli et al. by analysing Broad Copy Number Abnormalities (BCNAs)(aberration in at least 25% of chromosome length) in 46 colon samples and 26 normal mucosae an high frequency of gains (chr20q, chr8q, chr18, chr7q, chrXq, chr9q) and losses (18q, 8p, 17p, 15q) were found (Condorelli et al., 2018). The main results concerning the association between up-regulated driver cancer genes and the gene dosage effects due to gains of chromosomal number copies.

1.8 Aneuploidies in Glioblastoma

Glioblastoma multiforme (GBM) is a common type of malignant cerebral tumour. Based on the study conducted by The Cancer Genome Atlas Research Network (TCGA) and published on the Cell Journal (2013) with the title "The somatic Genomic Landscape of Glioblastoma", common chromosomal amplifications in GBM involve the chromosomes 1, 4, 7 and 12 by analysing 543 CGH samples data and performing a GISTIC interrogation (Brennan et al., 2013). Glioblastomas is characterized by isocitrate dehydrogenase 1 (IDH1) mutation in 12% of GBM patients. The glioblastomas without IDH mutations are characterized by chromosomal aberrations, such as gain of chromosome 7 and loss of chromosome 10 (Brennan et al., 2013). The chromosome 7 is specific feature of GBM and undergoes to a gain of the entire chromosome 7 (7-gain) alone or in combination with a focal amplification (FA) located on the cytogenetic band 7p11.2 (7p-FA), a region including the EGFR gene (Brennan et al 2013). In addition to focal amplification (7-PA), another EGFR mutational condition is defined as Regional gain in comparison to the Euploid definition which does not show alteration in EGFR region. Moreover, focusing on the role of the EGFR and its mutation with or without the 7-gain alteration, its genomic alterations are correlated with the specific molecular subtype of GBM. According to Brennan et. al 2013 and Verhaak et al., 2010 classifications(Brennan et al., 2013; Verhaak et al., 2010), GBM is divided in 4 subtypes (Classical, Mesenchymal, Neural, Proneural). As a consequence of a gene dosage effect, the transcriptional upregulation of some specific genes located in the 7p-FA region or widespread to the entire chromosome 7 has been repeatedly observed in the two aberrations, respectively (Beroukhim et al 2007, Brennan et al 2013). Other gliomas (Low grade gliomas) are characterized by 1p loss and 19q gain correlating with IDH mutation(TCGA, 2015). The co-gain of chromosomes 19 and 20 is also linked to a more favourable prognosis validated by a 124 cohort aCGH samples (Geisenberger et al., 2015). It is interesting that in a special type of glioma, so called oligodendroglioma, it has been found a co-deletion 1p/19q as the result of an unbalanced translocation der(1;19) (q10;p10) and this aberration correlates with a better prognosis (Jenkins et al., 2006). Indeed, deletions of 1p and 19q are associated with tumors with oligodendroglial components.

The Epidermal Growth Factor Receptor (EGFR) gene, Cyclin-dependent kinase 6 (CDK6) gene and MET Proto-Oncogene Receptor Tyrosine Kinase (MET), which are known to be putative oncogenes or tumour suppressor genes, are located on chromosome 7.

2. AIM OF THE THESIS

The present work of thesis proposes a comprehensive cytogenetics, genomics, and transcriptomics analysis of broad copy number alterations (BCNA) in the large cohorts of 20.000 tumour samples divided 33 cancer types proposes in The Cancer Genome Atlas (TCGA) project. The aim is to investigate the cancer aneuploidies in order to understand their crucial role in the mechanisms of cancer initiation and progression.

Currently, the analysis and experiments reported in the present thesis are focused on Breast Invasive Carcinoma (BRCA), Colon Carcinoma (COAD) and Glioblastoma multiforme (GBM). The BRCA has been explored both bioinformatically and experimentally in-vitro, respectively while for the cohorts of COAD and GBM pre-analysis and post-analysis ending in biological functional analysis have been performed. For the rest of the TCGA cohort pre-analysis has been performed by applying the experimental model and the bioinformatics pipeline in R language refer to chr7, chr8, chr13, chr18 and chr20 already applied for COAD. This thesis shows only the results refer to BRCA, COAD and GBM.

The main goal of BRCA project regards the following steps 1) the exploration of the most frequent chromosomal aberrations as reported in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (https: //mitelmandatabase.isb-cgc.org/, accessed on 1 August 2020) which are the derivative chromosome der(1;16) (q10;p10), formed by the short arm of chr16 and the long arm of chr1 and the isochromosome 1q, i(1q), formed by two long arms of chr1. 2) classification of BRCA tumours in so-called 1,16 chromogroups by the combined analysis of cytogenetics (SNP-arrays), transcriptomics (RNA-seq) and mutational (WES) data in order to investigate the chromosomal aberrations and their biological role 3) identifying new driver genes that can represent novel molecular targets for therapy 4) validating the bioinformatics results and the 1,16 chromogroups features by exploiting the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort composed from about 2509 invasive breast cancer samples, paired with 548 samples of normal breast tissue 5) evaluating of the prognostic significance of chromosomal aberrations bearing the 1,16 chromogroups in the TCGA series and performing a comparison with the same cytogenetics groups obtained from METABRIC cohort 6) identification of suitable breast cancer cell culture bearing the chromosomal aberration der(1;16)(q10;p10) and the other 1.16 chromogroups features by investigating 978 Copy Number Variation (CNV) cell lines data (Affymetrix SNP6.0) with 37279 gene coverage and 1047 RNA-seq cell lines data (counts data) with 20213 gene coverage retrieved from Cell Model Passport. 7) In-vitro

validation of the results obtained from previously steps on, mainly, CAL-148 (bearing the der(1;16) aberration) , CAL-51 (not bearing the cytogenetics aberrations) and the control cell group (MCF7, MCF10, MDA-MB-231) by performing drug-treatment and siRNA interference.

Other minor questions about the prognostic role of ETV6-NRTK3 fusion protein in secretory breast carcinoma, the immunohistochemistry (IHC) verification of HER2 expression protein and the triplonegativity of hormone receptors in CAL148, CAL51 and MCF10A have been explored.

The main goal of COAD project regards 1) The explorations of the main broad copy number gains (BCNGs) in those chromosomes resulting to be the most aberrant in COAD (chr7, chr8, chr13, chr18 and chr20) by analysing 433 colon adenocarcinoma (COAD) samples (data obtained by Affymetrix SNP 6.0 array) and expression profiles of 480 COAD samples and 41 mucosal normal samples (n=521) retrieved from GDC Data Portal 2) analysis of those tumours showing Chromosomal INstability (CIN) (233 tumours) and MicroSatellite Instability (MSI) (60 tumours) 3) Differential Expression Analysis of those genes up-regulated (PositiveT) and downregulated (NegativeT) in comparison to mucosal normal samples and those genes over-expressed (OverT) and down-expressed (DownT) in comparison to control tumours groups not bearing that specific BCNG 4) Biological and functional analysis of driver genes associated to BCNG and significantly enriched among Gained Variant Enhancer Loci Transcript (VEL-T) (identified by high resolution H3K27acChIP-seq profiling) as well as those genes required for cancer cell fitness (Fitness-genes) (identified by genome-scale CRISPR-Cas9 screening) 5) Pathways analysis of Over-Positive T and Fitness-OverT.

The main goal of GBM refers to a massive analysis of the chromosome 7 gain in those tumours with a high percentage of this BCNG. A general investigation of the percentage of trisomy 7 in 33 TCGA cancer types has been performed by using SNP array data. In TCGA the percentages of those tumour types with a high frequency of chr7-trisomy (> 45 %) are: GBM (67%), followed by KIRP (55.47%), COAD CIN (46.90%), READ CIN (45.09%). The subsequent steps are 1) analysis of cytogenetic SNP array data (592 samples), RNA-seq data (174 samples in total subdivided in 128 primary tumour and 5 normal tissue) and whole exome sequencing (WES) data (345 samples) from The Cancer Genome Atlas (TGCA) 2) analysis of the role of the Epidermal Growth Factor Receptor (EGFR) gene in association with or without the trisomy7 in GBM 3) Subtype classification (Classical, Mesenchymal, Neural, Proneural according to (Brennan et al., 2013; Verhaak et al., 2010)) and their

association with EGFR mutational status, which are defined as Regional gain, Focal amplification and Euploid 4) Other goals are to evaluate the overall survival of the most common mutations in EGFR (EGFRvIII, A289D/T/V and G598V/A) as well as to evaluate the co-gain chr19/chr20 5) Gene ontologies analysis.

My thesis effort could be also considered a work of brainstorming to organize and analyse large amount of data. Nowadays Big Data in healthcare industries, research, clinical application, oncology and so on are fast growing and expanding more and more. The new technologies of next generation sequencing and their update represent a significant challenge in terms of ability to data management, data analysis and post-analysis. The technologies needed to perform massive bioinformatics analysis, such as RNA-seq, SNP-arrays, whole exome-sequencing and the more recent single cell-RNAseq (scRNA-seq) produce terabytes of data, which requires an informatic background and the ability to organize and manage the data as well as the adequate resource to process them.

3. MATERIALS AND METHODS

The present work of thesis is focused on three different tumours, Breast invasive carcinoma (BRCA), Colon Carcinoma (CC) and Glioblastoma Multiforme (GBM), for which it has been performed bioinformatics analysis with a common initial approach based on 1) Data Collection of transcriptomics (RNA-seq) and cytogenetics (SNP-array) data 2) Filtering and data organization in according to the proposed model 3) Statistical Analysis 4) Post analysis.

The breast cancer was further explored with the following steps 5) Data Collection and analysis of transcriptomics and cytogenetics data refer to the cell lines 6) Cell cultures 7) Primers design 8) Grow-curves in MTT assay 9) gamma-secretase treatments 10) Fine-tuning the gene-silencing method on breast culture.

To better understanding in deep each tumours approach, this section is organized in three dedicated sub-sections: Breast invasive carcinoma (BRCA), Colon Carcinoma (CC) and Glioblastoma Multiforme (GBM).

3.1 BREAST INVASIVE CARCINOMA (BRCA)

3.1.1 Data collection: transcriptomics, cytogenetics and mutational data

The data about Breast Invasive Carcinoma (BRCA) concerning The Cancer Genome Atlas Project, were downloaded from The Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov) (Tomczak et al., 2015; Weinstein et al., 2013). The RNA-seq data (HTseq counts) are 1222 BRCA-TCGA samples. The Whole Exome Sequencing (WES) data (in Mutation Annotation Format (MAF) file which was analysed by SomaticSniper algorithms (Larson et al., 2012)) are 976 samples. The cytogenetic data (Affymetrix SNP 6.0 arrays) were downloaded from cBioPortal for cancer genomics (https://www.cbioportal.org) (Cerami et al., 2012; Gao et al., 2013) for a total of 1084 samples. The data were associated to the clinical, survival and classification data.

3.1.2 Pre-filtering

The RNA-seq data were filtered with the same method used for COAD samples. In particular only 1072 primary tumors (TCGA-##-##-01A only primary tumor as Sample Type and only unique Sample ID) samples and 99 normal mammary tissue samples (only TCGA-##-####-11A Sample ID) were selected. RNA-seq and SNP-arrays were matched together by using the corresponding sample id. 1058 tumour samples are shared between RNA-seq and SNP-arrays and 946 samples are shared among RNA-seq, SNP-arrays and WES-seq technologies (corresponding case ids were used). The Stable ensembl IDs (i.e. ENSG###) were matched with the corresponding gene name and additional annotation by using BioMart (Kinsella et al., 2011) and GRCh38.p13 genome version (Ensembl Release 99; January 2020). All deprecated genes between GRCh37 and GRCh38 genome versions were not considered.

3.1.3 Statistical Analysis: Tools and Approach

In comparison to the COAD-analysis described in par. 3.1, in BRCA analysis it has been introduced background noise removing by excluding low count genes (zero values in 70% of samples). The initial number of genes in RNA-seq count files were 60483; after pre-filtering we obtained 35923 genes and after removing deprecated genes 35903 genes. The RNA-seq BRCA counts were organized in the so called 1,16-chromogroups as described in the result section and then normalized as previously described in par. 3.1.3 Statistical Analysis: Tools and Approach.

The fragments per kilobase per million reads (FPKM) were converted in Transcripts per Million (TPM) according to the standard procedures. TPM was used to compare gene level expression in 1,16 chromogroups.

3.1.4 Post-Analysis: Semi-supervised Hierarchical Clustering, Pathway analysis

The data were organized in 1,16 chromogruops and undergo to the hierarchical clustering algorithm (Eisen et al., 1998)[24] in the pheatmap (Kolde, 2012) by using R packages. The row values were centred and scaled. The unweighted pair group method with arithmetic mean (UP-GMA) was used as linkage agglomeration method to apply at Euclidean distance.

The pathways analysis to investige the biological significance of the our genes list was performed by Metascape (Zhou et al., 2019) and with the Gene set enrichment analysis (GSEA) (Subramanian et al., 2005). The WES-seq data were analyzed by using the Maftools

R Package (Mayakonda et al., 2018). The Venn diagrams analysis were performed by Interactive Venn free software (Heberle et al., 2015)

3.1.5 METABRIC: Molecular Taxonomy of Breast Cancer International Consortium.

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) is the largest cohort of breast carcinomas. In particular, METABRIC includes 2509 samples of breast carcinoma and their corresponding 548 samples of normal mammary tissue. METABRIC study has explored breast cancer from transcriptomics and cytogenetics points of view. In addition, it focused on molecular-based taxonomy of breast cancer with the determination of 10 integrative cluster subtypes. The cohort also consider the Molecular subtype(LumA and LumB) and the Histological subtypes (Ductal and Lobular) (https://ega-archive.org) (Pereira et al., 2016)(Rueda et al., 2019)(Curtis et al., 2012) (Russnes et al., 2017)

3.1.6 Data collection: CNV METABRIC

The large data cohort from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) was downloaded from cBioPortal for cancer genomics (https://www.cbioportal.org)(Cerami et al., 2012)(Gao et al., 2013). In particular, 2173 CNV samples was analysed.

3.1.7 Overall survival (OS)

The overall survival (OS), the Progression- free survival (PFS) and the Disease-free survival (DFS) data were evaluated by the Kaplan Meier curve in GraphPad Prism software version 8.0 (GraphPad Software, San Diego, CA, USA, www.graphpad.com). Log-Rank Test was used.

3.1.8 Translation of Bioinformatics results in Cell Model

In order to translate the 1,16 chromogroups in suitable cell models, an additional bioinformatics analysis was performed. Transcriptomics and cytogenetics data were retrieved from Cell Model Passport (https://cellmodelpassports.sanger.ac.uk/) (Version data 20191101) (Van Der Meer et al., 2019): 1) 978 Copy Number Variation (CNV) cell lines data (Affymetrix SNP6.0) with 37279 gene coverage; 2) 1047 RNA-seq cell lines data (counts data) with 20213 gene coverage.

Only the cell lines matchable with the two technologies and referred exclusively to breast cancer were considered (52 Cell Lines)

The Cell Lines CNV Data were classified according to the classification in chromogroups 1,16 proposed in Privitera et al, 2021. The main 1,16 chromogroups are: 1) Group A, bearing 1q-gain and 16q-loss, 2) Group B, bearing 1q-gain/1p-loss, 3) Group C, bearing 1q-gain and normal chr16, 4) Group D, bearing 16q-loss and normal chr1, 5) Control (CTRL) No aberrations in chr1 and chr16.

The groups so composed were then analysed to individuate the differential genes (DEGs) by used the bioinformatic pipeline previously described in this section

.

3.1.9 Cell Model and Cell maintenance

The Cell Lines CNV Data were further used to investigate the aneuploidy assessment of cells in order to identify the breast cell model bearing the derivate(1;16) (CAL-148) and a control cell model (CAL-51) not bearing the aneuploidy.

To compare the cell models proposed with these two cell lines, CAL-148 e CAL-51, other three additional cell lines were used: MCF10A, MCF7 and MDA-MB-231.

The **Table 1** shows the karyotype and the chromosome 1 and 16 status of MCF10A, CAL-51, CAL-148, MDA-MB-231 and MCF7.

The cells maintenance has taken place at the laboratory of Cell Culture in the Department of Biomedical and Biotechnological Sciences - Section of Medical Biochemistry c/o Biological Tower, Via Santa Sofia n.89, Catania – Sicily – Italy.

Cell-lines	1,16 chromogrups Classification				Repositories and literature	
	1p	1q	16p	16q	Karyotype	References
MCF10A					The karyotype is stable and near-diploid, these cells are nonetheless cytogenetically abnormal. These cells harbor a deletion of the locus containing p16 and p14ARF, as well as amplification of MYC. Notably, MCF-10A cells express wild-type p53. MCF-10A cells are negative for estrogen receptor (ER). MCF-10A cells are almost diploid cells with several chromosomal abnormalities. MCF-10A cells have a reciprocal t(3;9)(p14;p21) translocation and an unbalanced translocation from chromosome 5q to the derivative 9. Loss of the p16 locus associated with t(3;9) translocation is believed to contribute to the immortalization of MCF-10A cells. Also, a t(6;19)(p25;q12) rearrangement can be seen in MCF-10A cells. Trisomies of chromosomes 20 and 16, and derivative of chromosome 8.	Debnath et al., 2003 Cowell JK et al., 2005 Zientek-Targosz, H et al., 2008 Brenner AJ et al., 1995
CAL-51	disomic	disomic	disomic	disomic	(ISCN notation) 46(45–46), XX Human near-diploid karyotype with 28% tetraploidy - 46<2n>XX, no consistent abnormality detected - rare example of tumor cell line with normal karyotype - resembles published karyotype	(Davidson et al., 2000) https://www.creative-bioarray.com
CAL-148	loss	gain	gain	loss	Human near-diploid karyotype with 16% polyploidy - 45-48<2n>XX, +1, +7, -8, -13, +20, add(1)(p36), der(1;16)(q10;p10), del(2)(q11q2?2), del(13)(q1?1q1?3), der(16)t(1;16)(q12-21 - p11) - resembles originator's unpublished karyotype	https://www.creative-bioarray.com
MDA-MB-231	loss	disomic	loss	loss	The cell line is aneuploid female (modal number = 64, range = 52 to 68), with chromosome counts in the near-triploid range. Normal chromosomes N8 and N15 were absent. Most of the chromosomes of the normal complement are disomic, but chromosomes No. 2, 6, 14, 15, 18, 19, and X are monosomic, whereas chromosomes No. 4, 5, 17, and 20 are trisomic. Chromosome No. 8 is missing in 4 of the 6 metaphases examined and chromosome No. 1 is trisomic in 4 out of 6 meta- phases. An 11q marker, an isochromosome for llq,i(11) (q ter---~cen--~q ter), is present.	Copyright 2016 ATCC All rights Reserved (Satya-Prakash et al., 1981)
MCF7	loss	40%loss 50%gain	47%loss 48%gain	57%loss 42%gain	65 (61–72).1×1, der(1)t(X;1), 2×2, der(2)t(2;3)(q36;?), 3×2, del(3)(p?), der(3)t(3;6), 4×3, 5×3, der(5)t(5;13)(p12?;q22), 6×1, der(6)t(6;7), der(6)t(3;6), 7×1, der(7)t(1;19;7;6), der(7)t(7;19;7), 8×1, der(?)t(16;11;8;11;3), der(8)t(8;15), der(8?)t(8;12), der(8)t(8;16), 9×2, der(?)t(6;20;9;3), der(9)t(8;9), 10×2, der(10)t(7;10)(?;q22), 11×2, der(?)t(11;1;17;19;17), 12×3, der(19)t(12;19), del(13)(q?), der(13)t(13;15;11;16), der(13)t(13;14), der(13)t(13;16), 14×2, der(?)t(7;14), 15×2, der(16)t(15;16), 16×2, 17×1, der(17)t(3;17), 18×2, der(18)t(18;21), 19×1, der(19)t(7;19), 20×1, der(?)t(17;1;19;17;20), der(20)t(3;20;1;20;1;20), der(20)t(20;1;21), 21×3, der(22)t(7;22), der(22)t(16;22), X×2 cp[6] **Cytogenetic Analysis:** modal number = 82; range = 66 to 87. The stemline chromosome numbers ranged from hypertriploidy to hypotetraploidy, with the 2S component occurring at 1%. There were 29 to 34 marker chromosomes per S metaphase; 24 to 28 markers occurred in at least 30% of cells, and generally one large submetacentric (M1) and 3 large subtelocentric (M2, M3, and M4) markers were recognizable in over 80% of metaphases. No DM were detected. Chromosome 20 was nullisomic and X was disomic.	(Davidson et al., 2000) Copyright 2016 ATCC All rights Reserved

Table 1. Karyotype of the cell lines.

CAL-148 (DMSZ No: ACC 460, breast carcinoma) cell line was purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures. The cytogenetic features describes a karyotype human near-diploid with 16% polyploidy - 45-48<2n>XX, +1, +7, -8, -13, +20, add(1)(p36), der(1;16)(q10;p10), del(2)(q11q2?2), del(13)(q1?1q1?3), der(16)t(1;16)(q12-21 - p11) (see **figure 8**)

CAL-148 were cultured in Dulbecco's Medium 4.5g/dL Glucose w/Glutamax (Gibco®) supplemented with:

- 20% fetal bovine serum (FBS; Cat. No. 10270-106; Life Technologies, Monza, Italy).
- 1% of 1/1 penicillin–streptomycin (Cat. No 15140- 122; Life Technologies).
- 10ng/ml Epidermal Growth Factor (EGF) (Cat. No E9644, Sigma-Aldrich, USA)

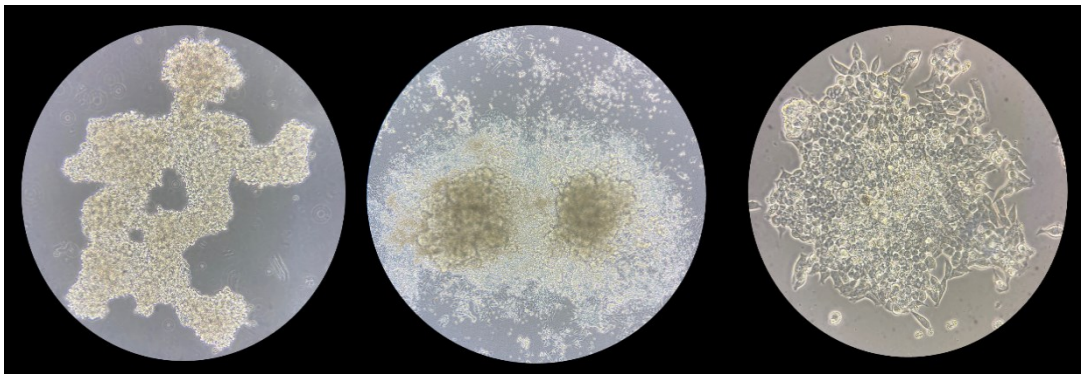


Figure 8: CAL148 (magnification 100x-400X) The images were obtained in our laboratory with an inverted microscope (Olympus CKX53). Typical clumps. No monolayer

CAL-51(DMSZ No: ACC 302, breast carcinoma) cell line was purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures. The cytogenetic features describe a karyotype with 28% tetraploidy - 46<2n>XX, no consistent abnormality detected - rare example of tumor cell line with normal karyotype (see **Figure 9**).

CAL-51 were cultured in Dulbecco's Medium 4.5g/dL Glucose w/Glutamax (Gibco®) supplemented with:

- 10% fetal bovine serum (FBS; Cat. No. 10270-106; Life Technologies, Monza, Italy).
- 1% of 1/1 penicillin–streptomycin (Cat. No 15140- 122; Life Technologies).

The cell cultures were grown in flasks (25 cm2) and incubated at 37 °C in humidified atmosphere with 5% of CO2 and 95% of air.

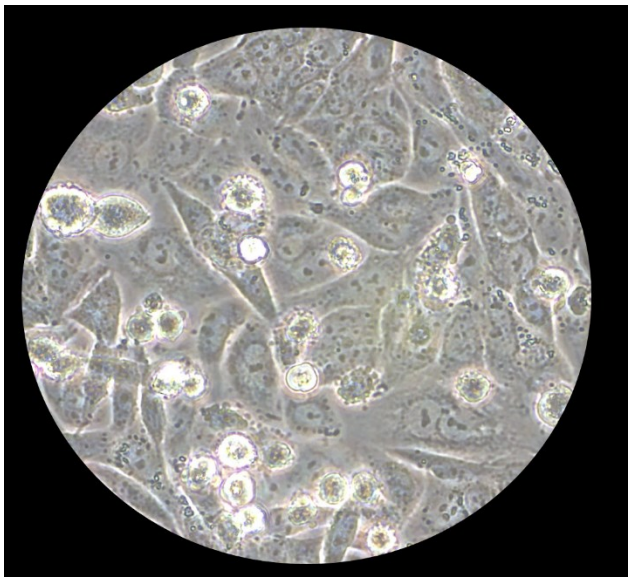


Figure 9: CAL51 (magnification 400x) The images were obtained in our laboratory with an inverted microscope (Olympus CKX53).

MCF10A (ATCC No: CRL-10317, normal'' breast epithelial cell line, Fibrocystic Disease) cell line was provided by prof. Luca Lanzanó and dott.ssa Morgana D´Amico from the Department of Physics and Astronomy "Ettore Majorana", University of Catania, Via S. Sofia 64, 95123, Catania, Italy (see **Figure 10**).

MCF10A were cultured in Dulbecco's Modified Eagle´s F12 w/15nM HEPES and sodium bicarbonate, w/o L-glutamine (Sigma®) supplemented with (final concentration):

- 5% Horse serum (HS; Cat. No. H1270; Sigma-Aldrich, USA).
- 10ug/mL Insulin (Cat.No., I-1882, 100-mg; Sigma-Aldrich, USA).
- 0.5ug/mL Hydrocortisone (Cat.No. H-0888, 1-g; Sigma-Aldrich, USA).
- 50ng/mL Cholera Toxin (Cat.No. C-8052, 2-mg; Sigma-Aldrich, USA).
- 1% of 1/1 penicillin–streptomycin (Cat. No 15140- 122; Life Technologies).
- 20ng/ml Epidermal Growth Factor (EGF) (Cat. No E9644, Sigma-Aldrich, USA)

The cell cultures were grown in flasks (25 cm2) and incubated at 37 °C in humidified atmosphere with 5% of CO2 and 95% of air.

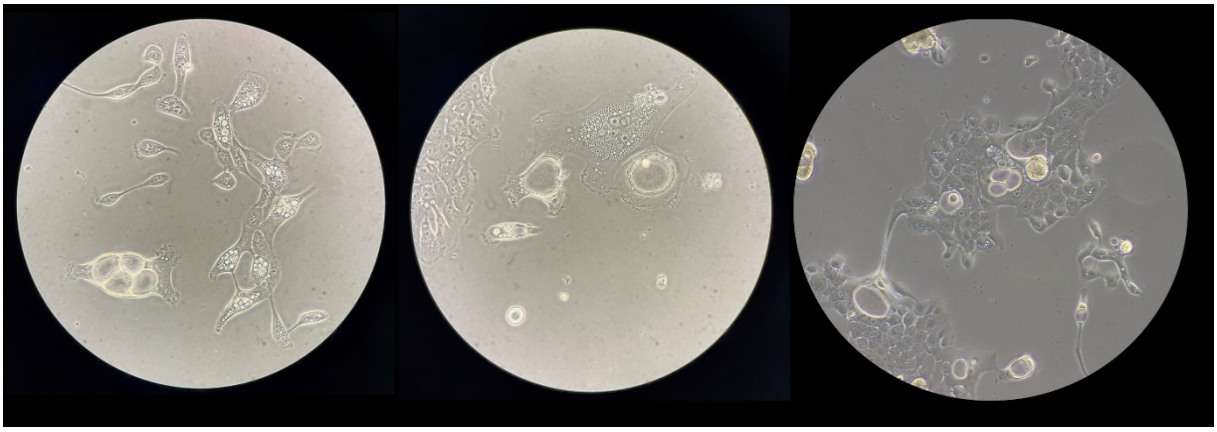


Figure 10: MCF10A (magnification 100x) The images were obtained in our laboratory with an inverted microscope (Olympus CKX53).

MCF7 (ATCC No: HTB-22, breast adenocarcinoma) cell line was purchased from the American Type Culture Collection (Rockville, MD, USA) (see **Figure 11**).

MCF7 were cultured in Dulbecco's Medium 4.5g/dL Glucose w/Glutamax (Gibco®) supplemented with:

- 10% fetal bovine serum (FBS; Cat. No. 10270-106; Life Technologies, Monza, Italy).
- 1% of 1/1 penicillin–streptomycin (Cat. No 15140- 122; Life Technologies).

The cell cultures were grown in flasks (25 cm2) and incubated at 37 °C in humidified atmosphere with 5% of CO2 and 95% of air.

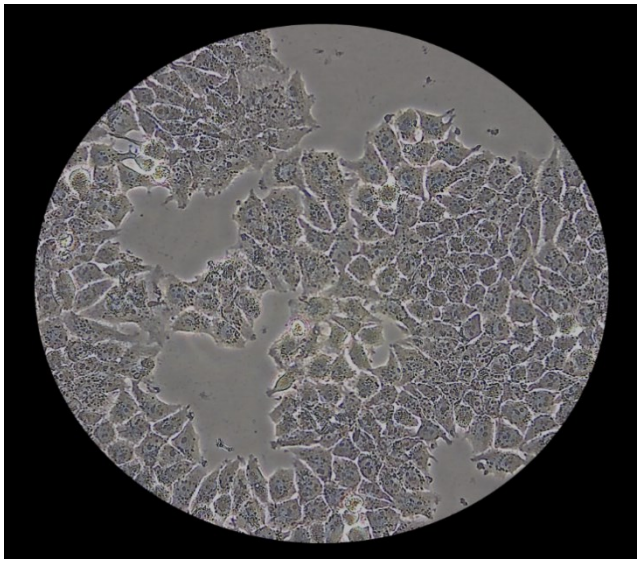


Figure 11: MCF7 (magnification 400x) The images were obtained in our laboratory with an inverted microscope (Olympus CKX53).

MDA-MB-231 (ATCC No: HTB-26, Breast Adenocarcinoma) cell line was purchased from the American Type Culture Collection (Rockville, MD, USA) (see **Figure 12**).

MDA-MB-231 were cultured in Dulbecco's Modified Eagle´s F12 w/o HEPES and L-glutamine (Gibco®) supplemented with:

- 10% fetal bovine serum (FBS; Cat. No. 10270-106; Life Technologies, Monza, Italy).

- 1% of 1/1 penicillin–streptomycin (Cat. No 15140- 122; Life Technologies).
- 1% of Glutamine
- 1 % EsseMEM Non-Essential Amino Acids Solution (100X) (Cat.No 11140050, Thermofisher)
- The cell cultures were grown in flasks (25 cm2) and incubated at 37 °C in humidified atmosphere with 5% of CO2 and 95% of air.
- FIGURA MDA-MB-231

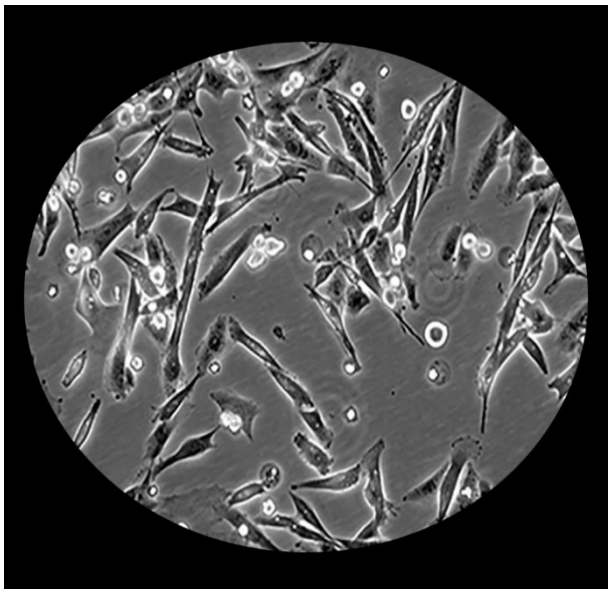


Figure 12: MDA-MB-231 (magnification 400x) www.mingzhoubio.com

3.1.10 Cell growth curve

The evaluation of a growth curve was useful to investigate the characteristics (i.e. population doubling time) of CAL-148, CAL-51, MCF7 with or without Epidermal Growth Factor (EGF). Four time-points (24h, 48h, 72h, 144h) and four cell concentration (2000, 4000, 8000, 16000 cells/well) were evaluated.

3.1.11 Cell Proliferation Assay

The evaluation of cell proliferation was performed with VybrantÆ MTT Cell Proliferation Assay Kit (Cat. No. V-13154, Themorfisher) according to the manufacturer's instructions.
The MTT solution was prepared by adding 1mL of sterile PBS 1X to one 5 mg vial of MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) and then distributed (20 µL) in each well (96well plate). Incubation for 2h at 37 °C. The purple formazan compound was solubilized by adding 200 µL of dimethyl sulfoxide-d6 (DMSO).

The absorbance is evaluated inside the Eppendorf 96 Flat Bottom Clear Polystyrene tissue-culture treated Cell Culture by using PlateReader AF2200. Wavelength at 600nm.

3.1.12 Primer design

Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and PrimerBank (Spandidos et al., 2008) (Spandidos et al., 2009) (X. Wang & Seed, 2003) online tools were used to primer design. The chosen criteria were:

1) PCR product size Min 70nt and Max 300nt,
2) Primer melting temperatures (Tm) Min = 58 °C, Opt = 59 °C, Max = 60 °C, Tm difference =1°C;
3) Exon junction span: primer must be an exon-exon span junctions and Primer pair must be separated by at least one intron on the corresponding genomic DNA;
4) Primer Size: Min = 18nt, Opt = 20nt, Max = 25nt;
5) Primer GC content (%): 40-60% but for *PYGO* 20-60%.

Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) was used to check the distance among exons and the exactly size among introns.. The primer design was performed for PYGO2, APH1A, NCSTN, PSEN2, CDH1 and BCL9 (**Table 2**). The primers have been ordered from Eurofins Genomics (https://eurofinsgenomics.eu/).

RefSeq mRNAs numbers	Transcript Variants	Gene Name	Right Primer	Left Primer	Amplicone Size (bp)	Chr arm	N exons	Intron size
NM_138300.4_New	//	**PYGO2**	AGTCCAGAAA AGAAGCGAAGG	CCGAAGTCATC TTCAAAAGGGT	131	1q21.3	3	2F_3R(1130bp)
NM_138300.4_Old	//	**PYGO2**	CCGGTCTGCA AATGAAGAGT	GTGATCCACCA TGGGAGTTG	113	1q21.3	3	1-2(447bp) 2F_3R 1130bp)
NM_015331.3	**Variant 1**	**NCSTN**	GGCAATGGTT TGGCTTATGAAG	TGGCACATAGT GGGAAGGTTG	142	1q23.1	17	5F 6R(971bp)
NM_001349729.2	Variant 5							
NM_001290184.2	Variant 2							
NM_000447.3	Variant 1	**PSEN2**	ACTCATCTG CCATGGTGTGG	TCCTCCTCTT CCTCCAGCTC	190	1q42.13	13	9-10 (1144bp) 10-11 (381/384 bp)
NM_012486.3	Variant 2							
NM_001077628.3	**Variant 1**	**APH1A**	GGTGGTTGGG AGTCACCTACT	GCGCTGAATA CTTCGGAGGG	148	1q21.3	7	5-6 (418 bp)
NM_016022.4	Variant 2							
NM_001243771.2	Variant 3							
NM_001243772.2	Variant 4							
NM_004360.5	**Variant 1**	**CDH1**	AAGTGTCCGA GGACTTTGGC	AATCCTCCCT GTCCAGCTCA	177	16q22.1	16	10-11 (3520bp)
NM_001317184.2	Variant 2							
NM_001317185.2	Variant 3							
NM_004326.4_New	//	**BCL9**	GCCAAAGTGG TGTACGTGTT	GATATCTGTGTGT TCAGAGGCG	155	1q21.1	10	6F_8R (1186bp)
NM_004326.4_Old	//	**BCL9**	CCCCATCAAAT GCTACAGCC	TTTCAACCTGG CCCTTCAAAA	165	1q21.1	10	6-7(1186bp)

Table 2: Primers features

3.1.13 RNA extraction

Starting from ~ 2 x 10^6 cells, total RNA was extracted from CAL-148, CAL-51, MCF7, MDA-MB-231, MCF10A cell lines, using RNeasy Mini Kit (Cat. No. 74104; Qiagen, Milan, Italy), according to the manufacturer's instructions. The RNA was eluted in 30 µL of RNAase Free Water.

The RNA extraction was also performed by using TRIzol™ Reagent (Cat. No. 15596026 and 15596018; Invitrogen, USA). Starting cell concentration of 2 x 10^6 cells. Cell lines: CAL-51 and MCF10A. The RNA was eluted in 30 µL of RNAase Free Water.

The concentration and quality of the RNA were determined using an ND-1000 spectrophotometer (NanoDrop, Thermo Scientific, Pero, Italy) (**Table 3**).

	RNeasy Mini Kit				**TRIzol™ Reagent**			
Cell line	ng/µl	260/280	260/230	Date	ng/µl	260/280	260/230	Date
CAL-148 (+ EGF)	725,1	2.03	1.34	17/06/2021	Not Performed			
CAL-148 (- EGF)	690.9	2.00	2.06	17/06/2021				
CAL-148 (Direct Lysis)	587.9	2.00	2.00	17/06/2021				
CAL-148 (Tryp + Lysis)	632.0	2.01	1.93	17/06/2021				
CAL-51	352,5	1.98	1.56	23/06/2021	1368.2	1.93	1.3	01/07/2021
MCF7	584.3	2.01	1.68	30/04/2021	Not Performed			
MDA-MB-231	1177.3	2.04	1.36	30/04/2021				
MCF10A	1166	2.00	1.79	23/12/2021	3246.7	1.74	1.55	26/07/2021

Table 3. RNA concentration of CAL148(+/- EGF), CAL-51, MCF7, MCF10A, MDA-MB-231 cell lines. Two method of RNAextration RNeasy Mini Kit and TRIzol™ Reagent

3.1.14 cDNA synthesis

cDNA synthesis of CAL-148, CAL-51, MCF7, MDA-MB-231 and MCF10A cell lines was also performed using Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit™ (cat. n.944404, Applied Biosystems), according to the manufacturer's protocol. The reactions were performed in a final volume of 20µL and 100ng/µL. Primers final concentration of 18pmol/µL.

3.1.15 Qualitative PCR

Qualitative PCR was performed using GeneAmp PCR System 9700 by Applied Biosystem (Applied Biosystems, Foster City, CA, USA).

The reaction (50 µL) was performed on 100 ng of cDNA from CAL-148, CAL-51, MCF7, MDA-MB-231 and MCF10A cell lines using Dream Taq PCR Master Mix (Thermo Fisher, cat n°K1082). The final concentration of each primer was of 18pmol/µL

The reaction was carried out using primer pairs described above and ACTB (Actin Beta, NM 001101.5) was used as internal control. Amplification conditions included a cycle of denaturation at 95 ° C for 3 minutes followed by 35 cycles of denaturation at 95 ° C for 30 seconds, annealing at 57 °C for 30 seconds and extension at 72°C for 1 minute, then a final extension takes place at 72° C for 10 minutes.

Following qualitative PCR, fragment sizes were confirmed carrying out with 2% agarose gel electrophoresis, using SYBR Safe staining (Invitrogen, cod. S33102). The electrophoresis run for 70 min at 90V.

3.1.16 Quantitative PCR

Quantitative PCR was performed by using StepOne Real-Time PCR Systems Applied Biosystem (Applied Biosystems, Foster City, CA, USA).

The reaction (20 µL) was performed on 1-2ug of RNA from CAL-148, CAL-51, MCF7,MDA-MB.231,MCF10A by using Kit Sybr Green PCR Master Mix (Cat.No. 4367659, Thermofisher) according to the manufacturer's protocol. The Primers final concentration of 18pmol/µL was used.

The reaction was carried out by using primer pairs described above and ACTB (Actin Beta, NM 001101.5) as internal control. Amplification conditions included a cycle of denaturation at 95 ° C for 10 minutes to activate the AmpliTaq Gold ® DNA Polymerase. followed by 40 cycles of denaturation at 95 ° C for 15 seconds, annealing at 58-60 °C for 1 minute. The melt curve stage at 95° C for 15 seconds, 60 °C for 1 minute and final step at 95° C for 15 seconds.

Gene expression is measured by the quantization of cDNA in relation to a calibrator sample serving as a physiological reference. The calibrator sample in each case is the cDNA from either the untreated cells or patients, or a specific tissue type. Results are also normalized to an endogenous control such as ACTB.

3.1.17 Immunocytochemistry (ICC)

The immunocytochemistry was only performed on CAL-148, CAL-51 and MCF10A in collaboration with prof.ssa Puzzo L and the dott.ssa Vecchio G.M from the Department of Medical and Surgical Sciences and Advanced Technologies, G.F. Ingrassia, Azienda Ospedaliero-Universitaria "Policlinico Vittorio Emanuele", Anatomic Pathology, School of Medicine, University of Catania, Italy

ErbB2 (HER-2) Monoclonal Antibody (CB11) (Cat. No. MA1-35720, ThermoFisher) was used as primary antibodies, with a dilution of 1:500.

Incubation with the secondary antibody (biotinylated anti-mouse immunoglobulins, Dako) and the peroxidase-conjugated streptavidin (Dako) was performed for 20 min at room temperature. Peroxidase activity was developed in the 3,3'diaminobenzidine (Sigma) substrate with 0.01% H_2O_2 for 5 min. Slides were counterstained with hematoxylin.

The slides so mounted were analysed at inverted microscope (Olympus CKX53).

3.1.18 Cell viability and Gamma secretase inhibitors.

The Gamma-secretase inhibitors were used to test the *in-silico* results showed in (Privitera et al., 2021). The drug used are:

- PF 3084014 hydrobromide (Cat.No. 5751, Tocris, Biotechne brand), Soluble to 10 mM in water with gentle warming
- DAPT (Cat.No. 2634, Tocris, Biotechne brand) Soluble to 100 mM in DMSO.

In **figure 13** the molecular structure is showed.

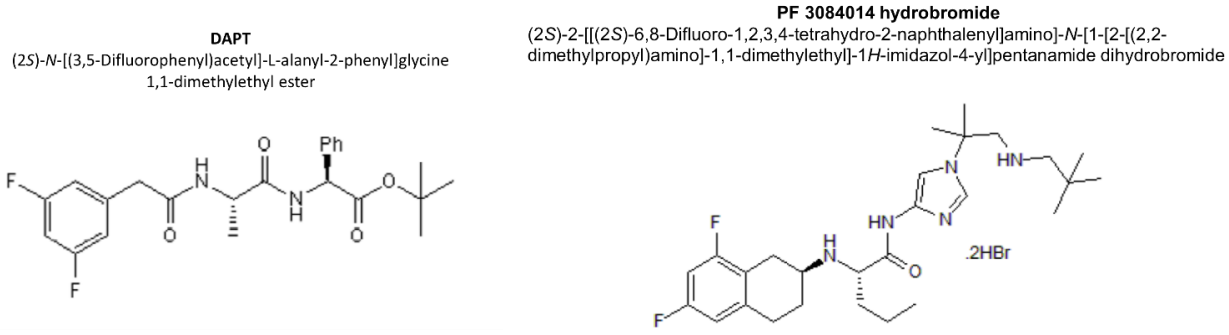


Figure 13: Molecular structure of Gamma-secretase inhibitors

Three different experiments were performed:

1) CAL-148, CAL-51 with or without Epidermal Growth Factor (EGF), three time-points (24h, 48h, 72h) and one cell plating density (4000 cells/well). Drugs concentrations (for both compounds) were 0.1 μM, 1uM, 10uM, 50uM.

2) MCF10A with EGF (the option without EGF was excluded because MCF10A cells are not able to survive without it), three time-points (24h, 48h, 72h) and one cell plating density (4000 cells/well). Drugs concentrations (for both compounds) were 0.1 μM, 1uM, 10uM, 50uM.

3) MCF10A and CAL148 with EGF, two time-points (48h, 72h) and one cell plating density (4000 cells/well). Drugs concentrations (for both compounds) were 0.1 μM, 1uM, 10uM, 50uM Both drugs were tested.

The efficacy of the treatment was evaluated with cell vitality assays kit. The absorbance was read by PlateReader AF2200 by Eppendorf. Wavelength of 600nm.

The half maximal inhibitory concentration (IC50) was calculated with GraphPad. The analysis settings are: 1) nonlinear-regression 2) Dose-response inhibition 3) log-inhibitor vs response (three parameters). The data were expressed as percentage of the control group (absorbance in the cells + vehicle (H_20 or DMSO only)).

3.1.19 siRNA transfection

The siRNAs were used to test the *in-silico* results showed in (Privitera et al., 2021) and they refer to the molecular component of the gamma-secretase complex: PSEN2, NCSTN, APH1A. Each siRNA (1pmol) was resuspended in 100µL of RNAase-free water to obtain 10uM according to the manufacturer's protocol.

- PSEN2 (Cat.No. 1027416, 1nmol siRNA, FlexiTube Gene Solution, Qiagen)

Product name	Cat.No	Target Sequencing
Hs_PSEN2_4	SI00009828	CAGCAGGTTTATCCAGATGAA
Hs_PSEN2_5	SI02623460	CAGGAGAGAAATGAGCCCATA
Hs_PSEN2_6	SI03038742	ACCATAGAAAGTGACGTGTTA
Hs_PSEN2_9	SI04380201	TTCATTGGATGCAGTTGTATA

- NCSNT (Cat.No.1027416, 1nmol siRNA, FlexiTube Gene Solution, Qiagen)

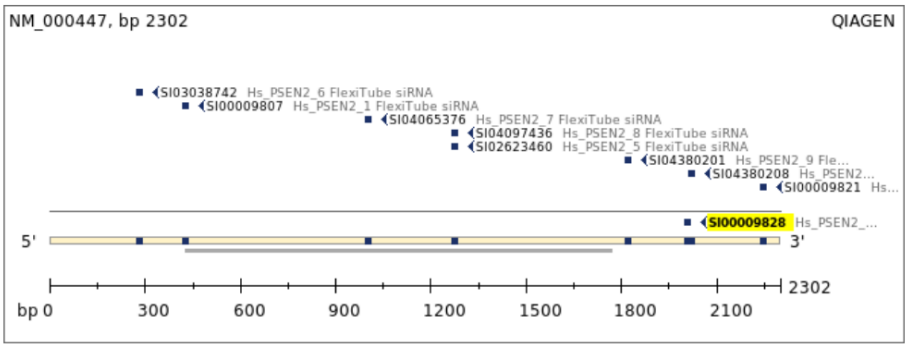
Product name	Cat.No	Target Sequencing
Hs NCSTN 8	SI02781198	AAGGTTTAATGTCAGGGTCAA
Hs NCSTN 7	SI02780953	TCAGATTGGGATTAACATAAA
Hs NCSTN 11	SI04439176	CAGGCAGGACCTAAGGTCCTA
Hs NCSTN 10	SI04439169	CTGAGAGCCGCTGGAAAGATA

- APH1A (Cat.No. 1027416, 1nmol siRNA, FlexiTube Gene Solution, Qiagen)

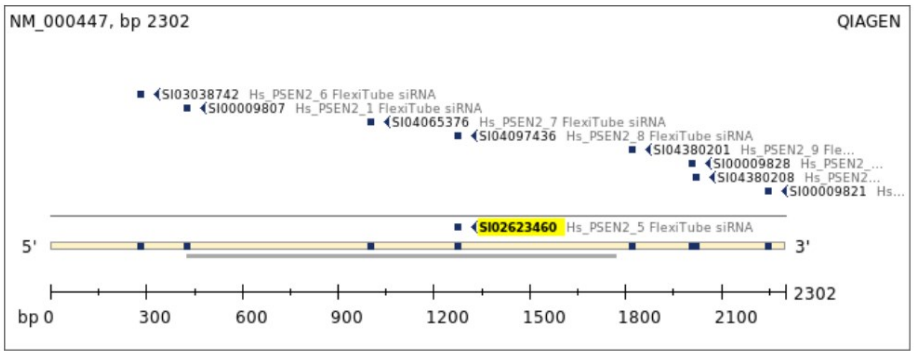
Product name	Cat.No	Target Sequencing
Hs_APH1A_1	SI02777516	TTGGTGTGATAAATACCCTAA
Hs APH1A 2	SI02777523	ATGAAATTAATGGAGGCTCAA
Hs APH1A 6	SI04386123	ATCGGGACTGACATTCCTGAA
Hs APH1A_7	SI04386130	AACTGGCATTACTGGAACTAA

A non-targeted siRNA sequence was designed according to non-human reference genes and Eurofins Genomics synthesized the siRNA. The sequence of positive-sense strand was 5′-[ACGUGACACGUUCGGAGAAUU]-3′ and antisense strand 5′-(AAUUCUCCGAACGUGUCACGU-3′).

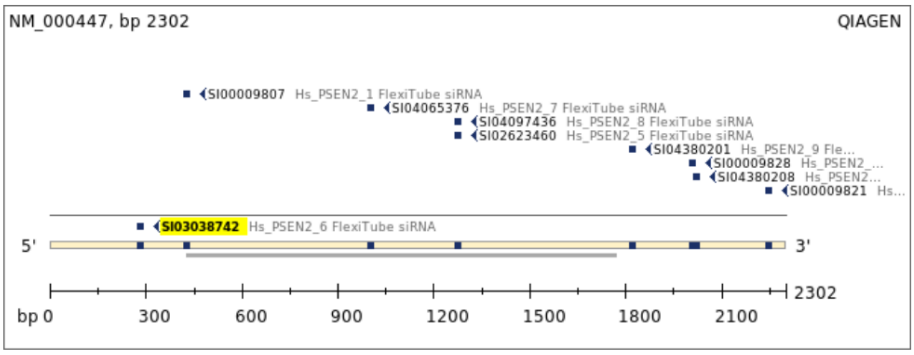
Figures **from 14 to 16** show the coding sequence where each siRNA span.



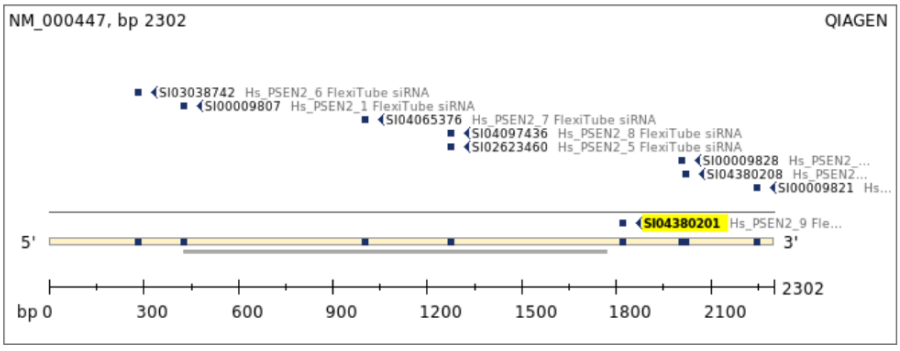
Hs_PSEN2_4
SI00009828



Hs_PSEN2_5
SI02623460



Hs_PSEN2_6
SI03038742



Hs_PSEN2_9
SI04380201

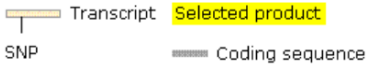
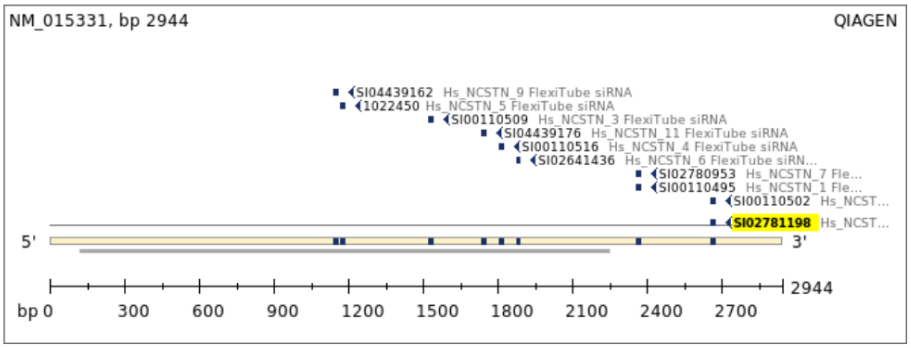
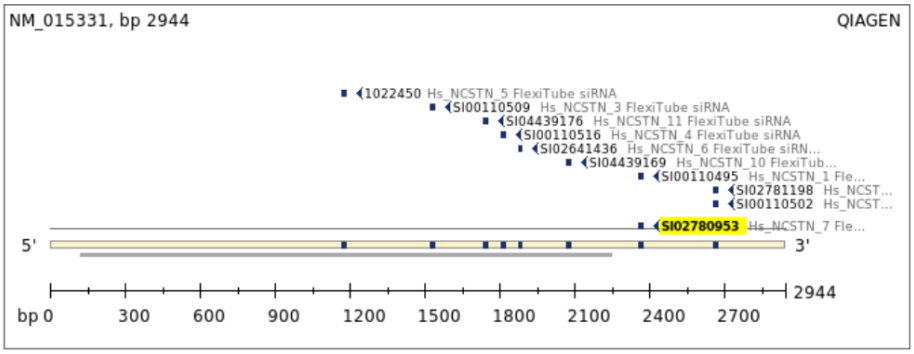


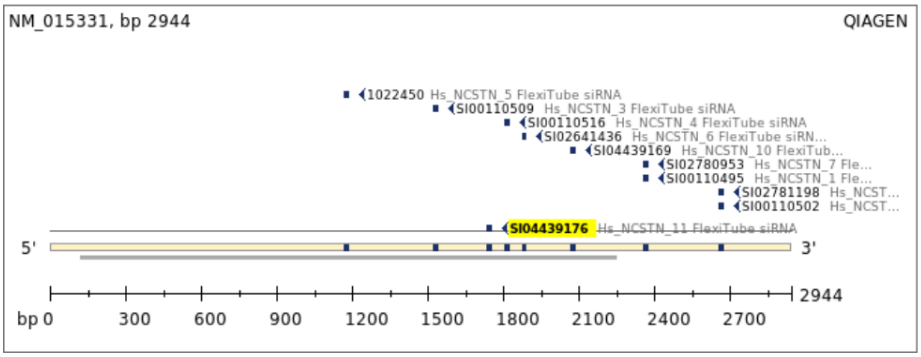
Figure 14: selected siRNAs spanning PSEN2 coding sequence (grey line) and transcripts sequence (yellow)



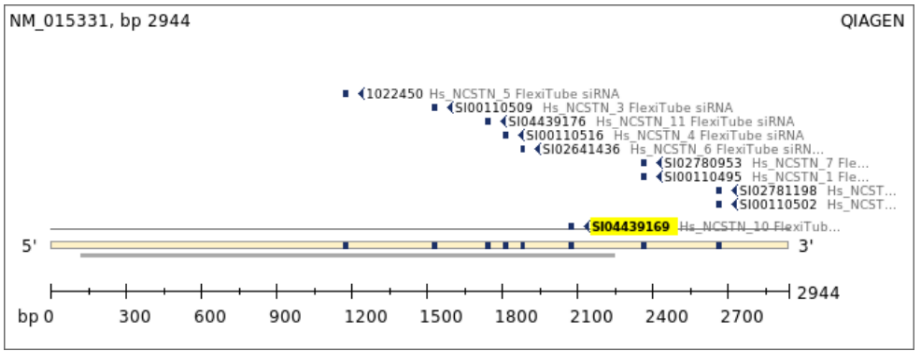
Hs NCSTN 8
SI02781198



Hs NCSTN 7
SI02780953



Hs NCSTN 11
SI04439176



Hs NCSTN 10
SI04439169

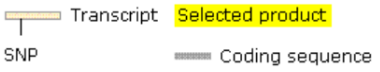
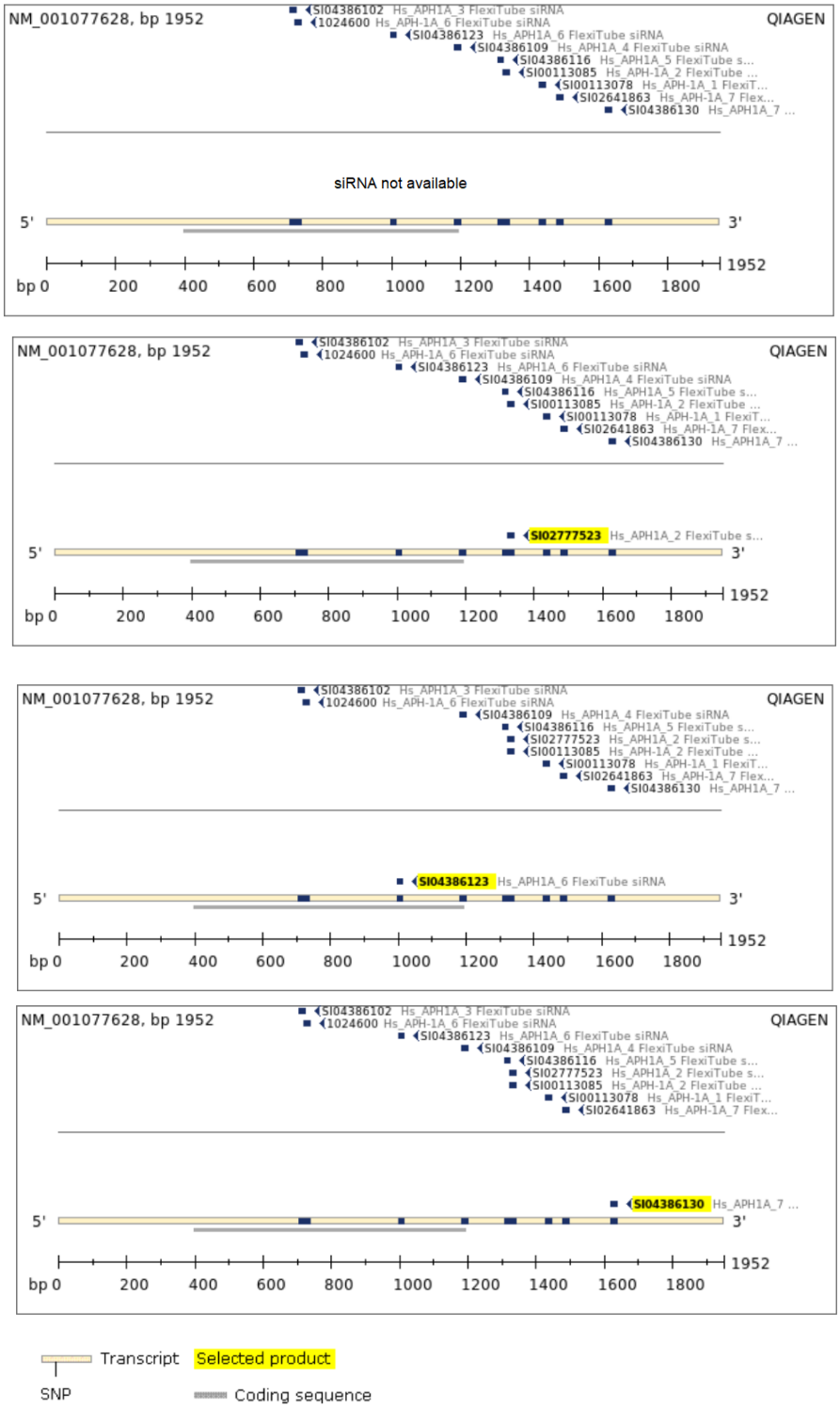


Figure 15: selected siRNAs spanning NCSTN coding sequence (grey line) and transcripts sequence (yellow)



Hs_APH1A_1
SI02777516

Hs APH1A 2
SI02777523

Hs APH1A 6
SI04386123

Hs APH1A_7
SI04386130

Figure 16: selected siRNAs spanning APH1A coding sequence (grey line) and transcripts sequence (yellow)

One experiment was performed by using MCF10A at 4000 Cells/well concentration and treated with PSEN2 siRNAs. three siRNA concentrations (10nM, 20nM, 40nM) and two time points (48h, 72h) were considered.

HiPerFect Transfection Reagent (Cat.No 301705, Qiagen) was used as permeabilizing reagent to form transfection complexes (siRNA+HiPerfect) by using "Transfection of Suspension Cell Lines" in 24-Well Plates according to the manufacturer's protocol.

From each well, RNA was extracted with RNeasy Mini Kit (Cat. No. 74104; Qiagen, Milan, Italy). The efficacy of the treatment and the cell response to gene silencing was evaluated in real time- one step. In **Table 4** are indicated the concentration(ng/µL) of the extracted RNAs and the calculation to obtain a final concentration of 50ng/µL in 20µL of final volume

RNA MCF10A 48h							
	Ci ng/µL	**260/280**	**260/230**	**Vf rex 20µL**	**Cf 50ng/µL**	**Vi (µL)**	**H20**
10nM Psen2 (1)	312.5	2.01	0.71	20	50	3.20	16.80
10nM Psen2 (2)	240.3	1.97	1.09	20	50	4.16	15.84
20nM Psen2 (1)	366.4	1.96	1.96	20	50	2.73	17.27
20nM Psen2 (2)	391.6	1.95	1.98	20	50	2.55	17.45
40nM Psen2 (1)	364	1.81	1.95	20	50	2.75	17.25
40nM Psen2 (2)	374.7	1.99	0.86	20	50	2.67	17.33
10nM non-targeted siRNA (1)	385.4	1.98	1.62	20	50	2.59	17.41
10nM non-targeted siRNA (2)	292.8	1.99	0.79	20	50	3.42	16.58
20nM non-targeted siRNA (1)	391.5	1.96	1.25	20	50	2.55	17.45
20nM non-targeted siRNA (2)	339.9	1.98	1.59	20	50	2.94	17.06
40nM non-targeted siRNA (1)	365.7	1.97	1.81	20	50	2.73	17.27
40nM non-targeted siRNA (2)	378.4	1.96	1.3	20	50	2.64	17.36
CTRL + HP (1)	424.2	1.91	1.45	20	50	2.36	17.64
CTRL + HP (2)	379.2	1.92	1.46	20	50	2.64	17.36
CTRL- (1)	333.7	2.02	0.83	20	50	3.00	17.00
CTRL- (2)	346.2	1.95	1.75	20	50	2.89	17.11
RNA MCF10A 72h							
	Ci ng/µL	**260/280**	**260/230**	**Vf rex 20µL**	**Cf 50ng/µL**	**Vi (µL)**	**H20**
10nM Psen2 (1)	395.8	1.95	1.95	20	50	2.53	17.47
10nM Psen2 (2)	447.8	1.93	1.13	20	50	2.23	17.77
20nM Psen2 (1)	429.3	1.93	1.95	20	50	2.33	17.67
20nM Psen2 (2)	446	1.92	1.1	20	50	2.24	17.76
40nM Psen2 (1)	433.2	1.94	1.1	20	50	2.31	17.69
40nM Psen2 (2)	434.2	1.93	1.41	20	50	2.30	17.70
10nM non-targeted siRNA (1)	432.4	1.93	1.93	20	50	2.31	17.69
10nM non-targeted siRNA (2)	412.9	1.99	1.48	20	50	2.42	17.58
20nM non-targeted siRNA (1)	405.9	1.93	1.92	20	50	2.46	17.54
20nM non-targeted siRNA (2)	411.6	1.94	1.45	20	50	2.43	17.57
40nM non-targeted siRNA (1)	385.8	2	0.85	20	50	2.59	17.41
40nM non-targeted siRNA (2)	396.9	1.94	1.94	20	50	2.52	17.48

CTRL + HP (1)	399.1	1.96	0.93	20	50	2.51	17.49
CTRL + HP (2)	361.9	1.96	1.85	20	50	2.76	17.24
CTRL- (1)	351.6	1.98	1.01	20	50	2.84	17.16
CTRL- (2)	381.4	1.98	1.36	20	50	2.62	17.38

Table 4: the concentration(ng/µL) of the extracted RNAs and the calculation to obtain a final concentration of 50ng/µL in 20µL of final volume

3.1.20 Real-time one step: MCF10A – PSEN2 silencing

Quantitative PCR-one-step was performed by using StepOne Real-Time PCR Systems Applied Biosystem (Applied Biosystems, Foster City, CA, USA).

The reaction (20 µL) was performed on 50ng/µL RNA from MCF10A by using KIT Quantinova SYBR Green RT-PCR KIT (Cat.No. 208152, Qiagen) according to the manufacturer's protocol (**Table 5**). The Primers final concentration of 6pmol/µL was used.

Components	**Volume X1 Reaction**	**X 24 => X 27 Reaction (3 Plus)**
2 X QN SYBR Green RT-PCR Master.Mix	10µL	270
QN ROX Reference Dye (Applied Biosystems cycler only) StepOne HighROX dye Cyclers	1:20 1 µL (20µL Vf)	25
QN SYBR Green RT-MIX	0.2 µL	5.4
10X primer mix Cf Primer Forward => 0.5 µM Cf Primer Reverse => 0.5 µM	1 µL F 1 µL R	27 µL F 27 µL R
QN IC RNA (optional)	NO	NO
RNA <= 200ng 100fg/reaction Cf Scelta = 50ng/µL	2uL	
H2O	4.8 µL	129.6
Volume Finale	20µL	540 µL

Table 5: Composition of reaction mix.

The reaction was carried out using PSEN2 primer pairs described above and ACTB (Actin Beta, NM 001101.5) as internal control. The reaction parameters are: 1) Reverse transcription for 10min at 50 °C, 2) PCR initial activation for 2min at 95 °C, 3)Denaturation for 5s at 95 °C, 4) Combined annealing/extension for 10s at 60 °C. Number of cycles 35-40.

3.2 COLON CARCINOMA (COAD)

3.2.1 Data collection: transcriptomics, cytogenetics data

The Colon adenocarcinoma (COAD) Transcriptomics Data (RNA-seq data) were obtained from The Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov) by selecting the Cancer Genome Atlas (TCGA), PanCancer program (Tomczak et al., 2015; Weinstein et al., 2013). We downloaded the RNA-seq data by obtaining 480 COAD samples and 41 mucosal normal samples (n = 521).

The cytogenetic results (Single-Nucleotide Polymorphism (SNP) Arrays data)obtained by Affymetrix SNP 6.0 arrays technology regarded 439 samples of COAD and were downloaded from cBioPortal for cancer genomics (https://www.cbioportal.org) (Cerami et al., 2012; Gao et al., 2013). From the entire amount of data, the only samples matched to TCGA-COAD RNA-Seq data were selected for further analysis (433 samples).

3.2.2 Pre-filtering

The RNA-seq data included different types of samples tissues identified by a number into the TCGA-Barcode. TCGA - ## - #### -01A = Primary tumour; -02A = Recurrent tumour; -01B-C; Primary Tumour vial B-C, -06A = Metastasis ; -11A Normal Mucosae Samples ; -10A =Blood Samples. The data were filtered by selecting only primary tumour tissue type (TCGA-##-##-01A only primary tumour as sample type and only unique sample ID) and removing the other types (-01B, -01C, -02A, -06A), thus providing 461 tumour samples and 41 mucosal normal samples. The Stable Ensembl gene IDs (i.e., ENSG###) were matched with the corresponding gene name and additional annotation by using BioMart (Kinsella et al., 2011) and Genome Reference Consortium Human Build 38.p13 genome version (GRCh38.p13)(Ensembl Release 99; January 2020). All deprecated genes between GRCh37 and GRCh38 genome assemblies were not considered.

3.2.3 Statistical Analysis: Tools and Approach

The data were analysed with adequate R packages to manipulate big amount of data. The raw RNA-seq data (counts data) were then normalized with the "trimmed mean of M-values (TMM)" method introduced by Robinson and Oshlack (Robinson & Oshlack, 2010). The TMM assumes that the majority part of genes is not differentially expressed and it does not consider gene length. The TMM is suitable in normalization of samples coming from

different tissues. The TMM is an implementation and a function in the used packages to perform differential expression analysis of transcripts. The packages are edgeR (McCarthy et al., 2012; Robinson et al., 2010) and compcodeR (Soneson, 2014). The outputs are a log-fold change (FC) between the comparison two chosen groups of analysis (i.e. tumour vs normal or tumour vs – control group). The p-values were adjusted for multiple comparisons according to the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). The log-fold changes were converted into linear fold-changes.

3.2.4 Post-Analysis: Grouping and Pathways Analysis

The data supplied with statistical significance and liner fold-change where further grouped according to their level of linear fold-change. For example, by selecting FCvsCTRL (linear fold-change (FC) between a selected group and the CTRL group) > 1.3, adjp < 0.05, the transcripts are enriched in specific chromosome arms. The number of transcripts in that part of chromosome is reported as percentage of the total number of transcripts and then normalized for the total number of transcripts. The parameter so described was previously introduced in (Condorelli et al., 2018; Condorelli et al., 2019).

NCDI of chromosomal region $n = \frac{x_n}{X_n} * \frac{1}{\sum_{i=1}^{T} \frac{x_i}{X_i}} * 100$

where xn is the number of transcripts belonging to a specific transcript class in the nth chromosomal region, Xn is the total number of transcripts in the nth chromosomal region, T is the total number of chromosomal regions subdividing the entire genome. The ratio between xn and Xn represents the chromosomal density of the transcript class.

The linear fold-change of the expression level and the adj-p-value < 0.05 are the specific parameters to identify the transcript class. In particular, the parameters are based on 4 different differential analyses

- FC1: Linear fold-changes obtained comparing all COAD samples vs. normal colonic mucosae
- FC2: linear fold-changes obtained comparing COAD samples bearing a specific arm-level CNA (Selected COAD group) to samples not bearing it (Control COAD group).
- FC3: linear fold-changes obtained comparing “Control COAD” group to normal colonic mucosae.
- FC4: linear fold-changes obtained comparing “Selected COAD” group to normal colonic mucosae.

Applying specific threshold to these parameters, different transcript classes are determined:

- **UpT** (FC2 > 1.3 and adj-pvalue < 0.05)
- **DownT** (FC2 < 1.3 and adj-pvalue < 0.05)
- **Over-UpT** (FC2 > 1.3 and adj-pvalue FC2 < 0.05 + FC3 >1 and adjpvalueFC3 < 0.05 + FC4 >1 and adjpvalueFC4 < 0.05)
- **Under-DownT** (FC2 < 1.3 and adj-pvalue FC2 < 0.05 + FC3 <1 and adjpvalueFC3 < 0.05 + FC4 <1 and adjpvalueFC4 < 0.05).

The data so organized are then submitted to pathway analysis by using the dedicated software Ingenuity Pathway Analysis (IPA) (QIAGEN Inc.).

3.3 GLIOBLASTOMA MULTIFORME (GBM)

3.3.1 Data collection: transcriptomics, cytogenetics and mutational data

Data about Glioblastoma (GBM) produced by The Cancer Genome Atlas (TCGA) study (Taylor AM et al. 2018), were downloaded from the online resource *The Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov)*(Tomczak et al., 2015; Weinstein et al., 2013). Two types of data were selected 1) the RNA-seq data (counts data) by selecting 144 GBM samples (only primary tumour) and 5 normal tissue and 2) the Whole Exome Sequencing (WES) data. The molecular cytogenetic analysis of 592 TCGA-GBM (Affymetrix SNP 6.0) (Taylor et al., 2018), were retrieved from *cBioPortal for cancer genomics* (https://www.cbioportal.org)(Cerami et al., 2012; Gao et al., 2013)

3.3.2 Pre-filtering

The only primary tumours samples (RNA-seq counts and SNP-arrays data) were matched by using the corresponding sample id. The 128 tumour samples were in common between RNA-seq and SNP-arrays and 60 samples among RNA-seq, SNP-arrays and WES-seq technologies by using the corresponding case id. The annotation was matched with the corresponding stable Ensembl IDs (i.e. ENSG###) by using BioMart (Kinsella et al., 2011). The genome versions were GRCh37 and GRCh38.p13 genome version (Ensembl Release 99; January 2020). All deprecated genes were excluded.

3.3.3 Statistical analysis: RNAseq normalization, Tools and Approach

The gene counts showing a low count (zero values in at least 70 % of the samples) were excluded. The preliminary number of genes was 60483; subsequently the exclusion of low count genes, we obtained 37450 genes. The samples were organized in groups according to their cytogenetics features as well as the presence or absence of focal amplification of EGFR. The normalization was performed by using the Trimmed Mean of M-values (TMM) algorithm (Robinson & Oshlack, 2010) within the edgeR (McCarthy et al., 2012; Robinson et al., 2010) and compcodeR (Soneson, 2014) packages. The gene expression ratio is expressed in log-fold change as well as in linear fold-change. Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) were used to adjust the p-values for multiple comparisons. The Differentially expressed genes (DEGs) of the analysed cytogenetics groups were obtained thought the linear-fold change value which evaluates the expression level of the considered

cytogenetic group versus the CTRL group (group without chromosomal aberrations for chromosome 7).

Subsequently, the Normalized Chromosomal Distribution Index (NCDI) was calculated to evaluate the densities of each class of transcripts in a specific chromosomal, region according to the formula described in par. 3.1.4 Post-Analysis: Grouping and Pathways Analysis.

3.3.4 Post-Analysis: Semi-supervised Hierarchical Clustering, Pathway analysis and GO analysis

The GBM data already normalized in fragments per kilobase per million reads mapped (FPKM) were downloaded from The Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov). The FPKM values were then converted in Transcripts per Million (TPM) according to the standard conversion procedures.

The TPM values were used to perform a hierarchical clustering (Eisen et al., 1998) by using the pheatmap package (Kolde, 2012). The genes correspond to the values in the rows, while the samples correspond to the columns. The used data were normalized by using a modified z-score as indicated in formula (1).

$$\text{Modified z-score} = Z - score = \frac{\text{X} - X\ of\ CTRL}{\sigma\ of\ CTRL} \quad (1)$$

The agglomeration method was the unweighted pair group method with arithmetic mean (UPGMA) applied to Euclidean distance.

The pathways analysis and the biological significance of the analysed groups were performed by using the **Gene Set Enrichment Analysis** (GSEA) an available software GSEAv4.1.0 [build:27], www.gsea-msigdb.org/gsea/index.jsp (Subramanian et al., 2005)

In addition a gene ontology (GO) analysis was performed in two steps: 1) The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System (Mi et al., 2021) was used to categorize the gene list in Biological Function (BF), Molecular Process (MP) and Cell Component (CC) 2) GO results were elaborated by using GOPlot(Walter et al., 2015) R package.

The WES-seq data were analysed, summarized and annotated by using the Maftools R Package (Mayakonda et al., 2018). The Venn diagrams were generated by the Venn diagrams tool in Van de Peer Lab (http://bioinformatics.psb.ugent.be/webtools/Venn/). The whole data organization was performed entirely in R Studio (RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com). The statistical analysis was performed both in R than with GraphPad Prism software version 8.0.

3.3.4 Overall survival (OS)

The overall survival (OS), the Progression- free survival (PFS) and the Disease-free survival (DFS) data were evaluated by the Kaplan Meier curve in GraphPad Prism software version 8.0 (GraphPad Software, San Diego, CA, USA, www.graphpad.com). Log-Rank Test was used.

4. RESULTS

4.1 BREAST INVASIVE CARCINOMA (BRCA)

In this work, we exploited the large amount of molecular cytogenetic data (single nucleotide polymorphism (SNP) array data) provided by The Cancer Genome Atlas (TCGA) study (http://cancergenome.nih.gov/ (accessed on 29 October 2019)) in order to generate groups of breast invasive carcinomas (here called 1,16-chromogroups) characterized by a pattern of arm-level somatic copy number aberrations congruent with the different cytogenetic abnormalities of chromosomes (chr) 1 and 16 (see introduction section 1.5). After the cytogenomic characterization of 1,16-chromogroups, three main types of analysis of differential gene expression among 1,16-chromogroups were performed: (I) involving all BRCA samples, (II) restricted to ductal adenocarcinomas belonging to the LumA subtype, and (III) focused on the differences between ductal and lobular invasive adenocarcinomas. A general schematic workflow of the bioinformatics analysis performed for the present thesis is shown in **Figure 16**. The **table 6** specifies the number of samples used for the bioinformatics analysis.

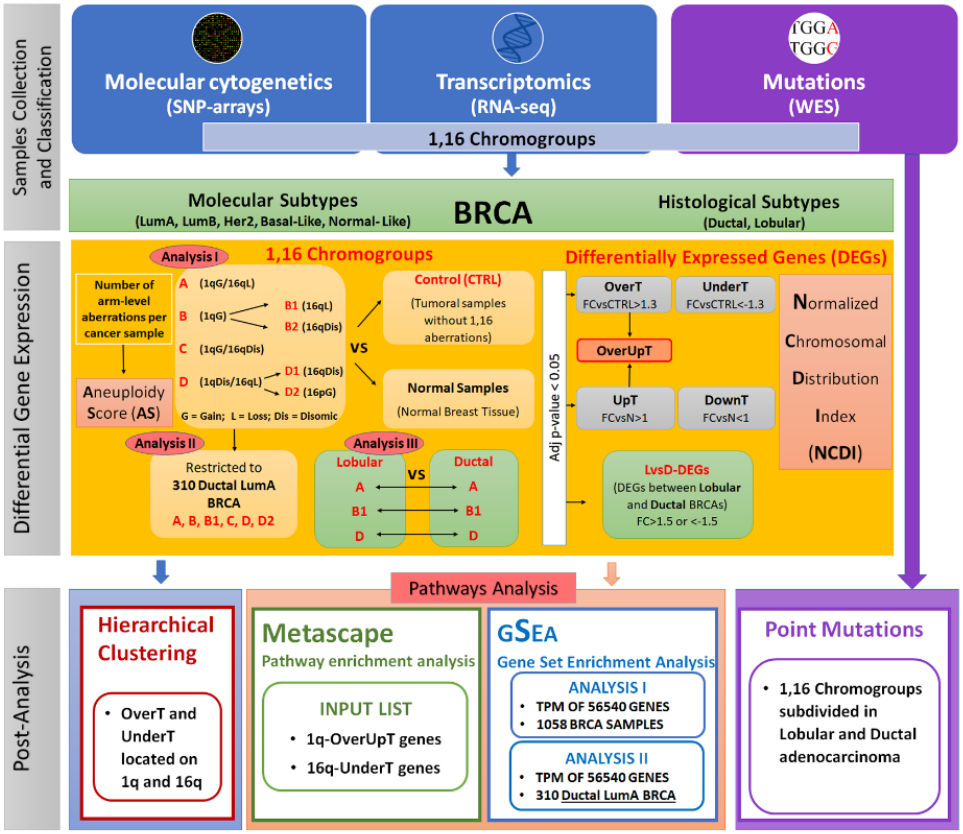


Figure 16: Schematic workflow of the bioinformatics analysis

Arm-level copy number				1,16-chromogroup name	Brief description of copy number aberrations in chr1 and chr16	Cytogenetic chromosome aberration inspiring copy number criteria	Number of samples		
1p	1q	16p	16q				SNP array	RNA-Seq	WES
D	G	G/D	L	**group A**	1q-gain and 16q-loss	der(1;16)	178	175	151
L	G	NF	NF	**group B**	1q-gain/1p-loss	i(1q)	171	165	151
L	G	G/D	L	**subgroup B1**	B with 16q-loss	i(1q)	101	98	98
L	G	D	D	**subgroup B2**	B w/o 16q-loss	i(1q)	18	17	17
NF	G	D	D	**group C**	1q-gain and normal chr16		90	89	85
D	D	G/D	L	**group D**	16q-loss and normal chr1		75	72	69
D	D	D	L	**subgroup D1**	D w/o 16p-gain	del(16q)	28	27	25
D	D	G	L	**subgroup D2**	D with 16p-gain		47	45	44
D	D	D	D	**Control (CTRL)**	No aberrations in chr1 and chr16		71	68	48

Table 6: Number of samples and arm-level copy number criteria for each 1,16-chromogroups. G: gain; L: loss; D: disomic; NF: no filter during selection; w/o: without.

4.1.1 Cytogenomics by SNP Array and 1,16-Chromogroups

Figure 17 shows the frequencies of chromosomal arm-level aberrations for the entire series of 1084 BRCA samples from TCGA. The gain of chr1q and the loss of chr16q were the most detected abnormalities followed by 17p-loss, 16p-gain, 8p-loss, 8qgain, 22-loss, 13q-loss, and 20q-gain.

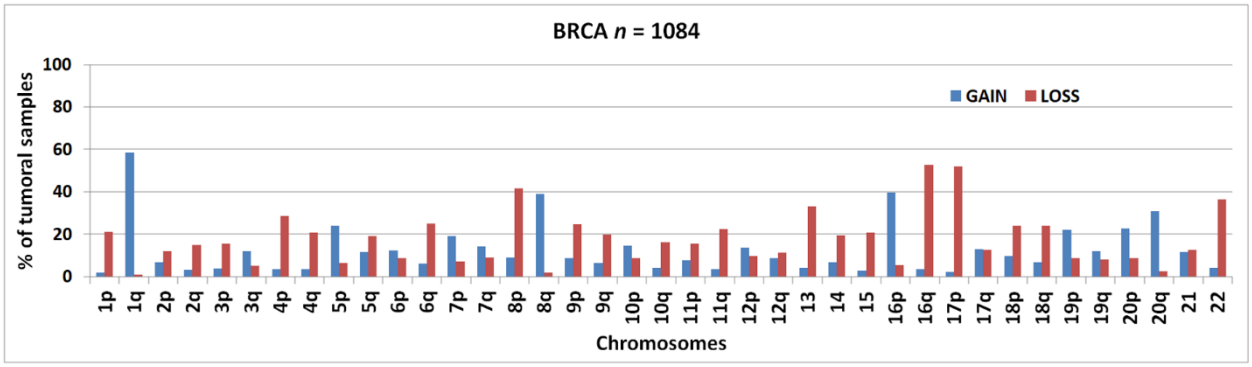


Figure 17: Percentage of samples bearing chromosomal arm-level gains or losses in the cohort of SNP array samples.

In order to study the transcriptional effects of aberrations of chr1 and chr16, alone or in combination, we defined different subgroups of breast cancer samples according to the copy number status of those two chromosomes (1,16-chromogroups). As reported in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (https: //mitelmandatabase.isb-cgc.org/, accessed on 1 August 2020), two common cytogenetics

abnormalities in breast cancer can underlie the gain of 1q: the derivative chromosome der(1;16) (q10;p10), formed by the short arm of chr16 and the long arm of chr1, as well as the isochromosome 1q, i(1q), formed by two long arms of chr1.

Considering the copy number changes frequently associated with the presence of der(1;16) (**Figure 6**), we selected 178 BRCA samples composing group A, as reported in **Table 6**. Though we could not formally exclude the presence of other cytogenetic abnormalities leading to the same pattern of arm-level copy number changes in chr1 and chr16, we assumed that der(1;16) is highly enriched in tumours of group A, relying on the fact that it is one of the most common 1q-aberration shown in conventional cytogenetics studies in breast cancer (Dutrillaux et al., 1990; Farabegoli et al., 2004; Kokalj-Vokac et al., 1993; N. Pandis et al., 1994; Nikos Pandis et al., 1992, 1995; Rye et al., 2015; Teixeira et al., 2002; Tsarouha et al., 1999).

Moreover, group A showed distinctive properties from another group bearing concomitant 1q-gain and 16q-loss (group B1). The i(1q) is another common aberration that produces a 1q-gain aberrations, and it was found to be as frequent as der(1;16). It is thought to be derived by an anomalous chromatid separation that leads to the generation of a chromosome formed by two 1q arms. Such aberration produces a 1q-gain associated with 1p-loss (**Figure 6**, right panel top). According to this pattern of copy number changes, we formed the so-called group B (n = 171). The groups A and B were disjoint because of the different condition established for Chr1p (disomic in group A or lost in group B). In group B, as a whole, we did not impose any criteria relative to the copy number status of chromosome 16 (**Table 6**), although a large fraction of samples of group B (67%) were found to bear a loss of 16q. However, we also formed a subset of group B, denominated subgroup B1, in which only samples showing a lost 16q and a gained or a disomic 16p were included (n = 101). Another subgroup of B, called B2, was formed by samples of group B bearing a disomic Chr16 (n = 18). Finally, a group C (n = 90) was formed by selecting samples bearing 1q-gain and disomic Chr16. Groups C and A were disjoint because of chr16 status, but group C was found to partially overlap with group B since it fully included subgroup B2.

In order to study the effects of the loss of chr16 not accompanied by aberrations of chr1, we also generated a group D (n = 75) characterized by normal Chr1 and 16q-loss. This group could be subdivided in two subgroups: the first one (n = 28), called subgroup D1, was formed by samples bearing 16q-loss and no arm-level copy number abnormalities in 16p, 1p, and 1q. Chr16 aberration in group D1 may correspond to the deletion of chr16q, del(16q), as observed in conventional cytogenetic studies (**Figure 6**). The second subgroup D (n = 47), called D2,

was characterized by normal chr1, 16q-loss, and 16p-gain. No clear correspondence to reported conventional cytogenetic abnormalities could be identified for this subgroup. Subgroups D1 and D2 were disjoint because of the difference in chr16p (disomic in subgroup D1 and lost in subgroup D2).

In summary, group A and subgroup B1 contained BRCA samples bearing concomitant 1q-gain and 16q-loss, group C and subgroup B2 were formed by BRCA samples bearing 1q-gain in the absence of aberrations of chromosome 16, and Group D and its subgroups D1/D2 were formed by samples bearing 16q-loss in the absence of aberrations of chr1 (**Table 6**). Finally, we formed a group containing cancer samples not bearing any arm-level aberrations in chr1 and chr16. In the context of the present analysis, the latter group played a special role and was denominated control (CTRL) cancer group for this reason (**Table 6**). The basic assumption was that CTRL tumours follow a different evolutionary pathway towards malignancy. Therefore, the analysis of differential gene expression (other chromogroups versus CTRL group), was used to generate lists of putative dosage-sensitive cancer driver genes associated with 1q-gain and/or 16q-loss.

Figure 18 shows the frequencies of arm-level aberrations in the above-defined 1,16-chromogroups. In accordance with the procedure followed for group formation, the main differences between the various groups involve chr1 and chr16. However, an increased frequency of 8q-gain in subgroups B1 and D2 and an increased frequency of losses of several other chromosomes in group B should be noted.

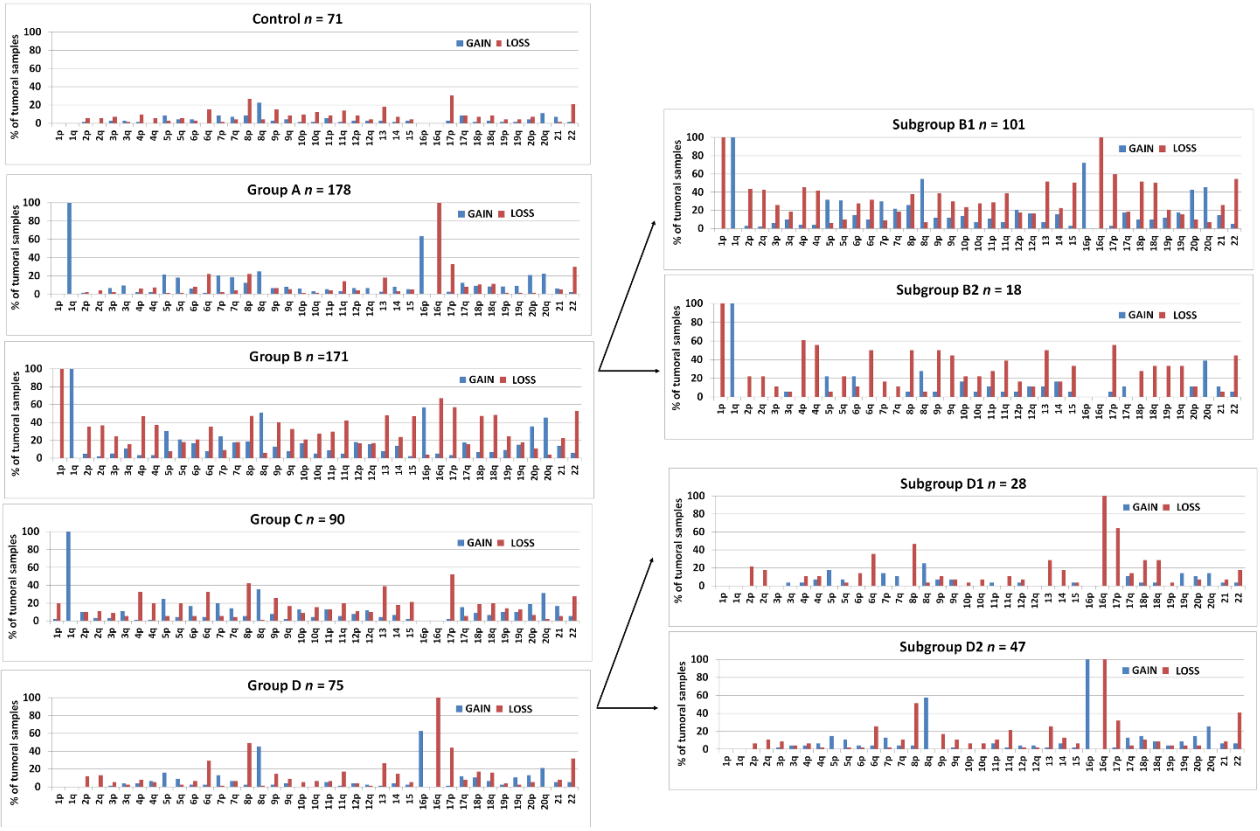


Figure 18. Bar graphs showing the percentage of samples bearing arm-level gains or losses in chromosomes 1–22 in the different 1,16-chromogroups (A, B, C, D, B1, B2, D1, D2, and control) of BRCA samples. The number of samples (n) in each group is reported in the corresponding graph.

Figure 19 shows the distribution of the aneuploidy score (AS) (i.e., the number of arm-level aberrations per cancer sample) for each 1,16-chromogroup. The control group showed the lowest averaged AS (mean ± SD: 4.46 ± 5.53; median: 2; interquartile range (IQR): 0–6). It is interesting that an enrichment of samples with a low AS was observed in Group A (mean ± SD: 8.24 ± 6.32; median: 6; IQR: 4–10) and Group D (mean ± SD: 7.88 ± 5.75; median: 6; IQR: 4–9), while an enrichment of samples with a high AS was observed in group B (mean ± SD: 18.30 ± 6.96; median: 19; IQR: 13–23) and subgroup B1 (mean ± SD: 19.49 ± 6.82; median: 21; IQR: 16.5–23.5). Indeed, the increased AS in groups B/B1 was due to the higher frequency of arm-level aberrations, mainly losses, in several different chromosomes (**Figure 18).** An intermediate value of AS was observed in subgroup B2 (mean ± SD: 13.88 ± 6.27; median: 14; IQR: 9.5–16.25) and group C (mean ± SD: 10.53 ± 7.04; median: 9.5; IQR: 4–16) (**Figure 19**).

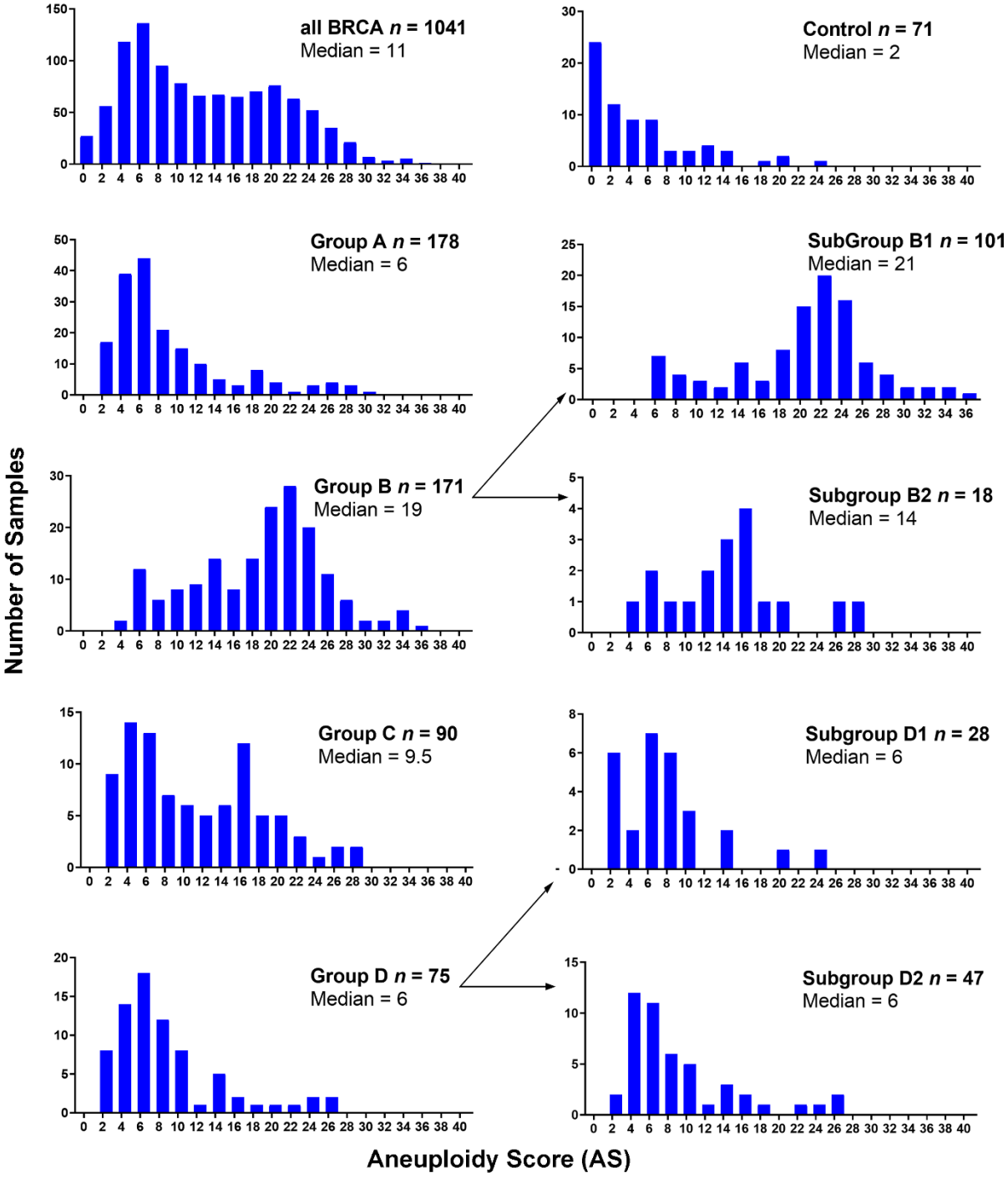


Figure 19. Distribution of the aneuploidy score (AS) for all BRCA samples and for each 1,16-chromogroup. Number of samples (y-axis) for different AS values (x-axis; bin equal to two) are shown. The total number of samples (n) and the median AS in each group is reported in the corresponding graph.

4.1.2 Transcriptomics in 1,16-Chromogroups

The differential expression analysis of transcript levels between different 1,16-chromogroups was performed by the edgeR package (McCarthy et al., 2012; Robinson et al., 2010) and expressed as the linear fold-change (FC) between a selected 1,16-chromogroup and the CTRL group (FCvsCTRL). The numbers of samples in each group are reported in **Table 6** and **Figure 20**. We called OverT (Overexpressed Transcripts) or UnderT (Underexpressed Transcipts) those transcripts expressing a value of the FCvsCTRL >1.3 or <-1.3, respectively. OverT and UnderT were selected at a false discovery rate adjusted *p*-value (adjp) of <0.05. As shown in **Figure 20**, the chromosomal distribution of OverT and UnderT genes (expressed as normalized chromosomal distribution index (NCDI); (Condorelli et al., 2019) was in agreement with the arm-level aberrations selected for each chromogroup (compare **Figures 18 and 20).** In order to easily identify the modifications of the chromosomal distribution of OverT and UnderT, the chromosomal distribution of all transcript-encoding genes (*n* = 56,540) is also reported in **Figure 20**. For instance, Group A showed an increased density of OverT in 1q and 16p and a decreased density in 16q, in agreement with the fact that copy number aberrations in those chromosomal arms were used as criteria for the formation of such group (**Figure 20**). A similar correlation could be observed in all other 1,16-chromogroups. Such correlation extended to chromosomal arm aberrations that were not primarily selected during formation of the 1,16-chromogroups, such as the increased frequency of 8q gain in subgroups B1 and D2. This was an expected result that is easily explained by the well-known "gene dosage transcriptional cis-effect" reported in several published studies (see references in (Condorelli et al., 2019; Condorelli et al., 2018).

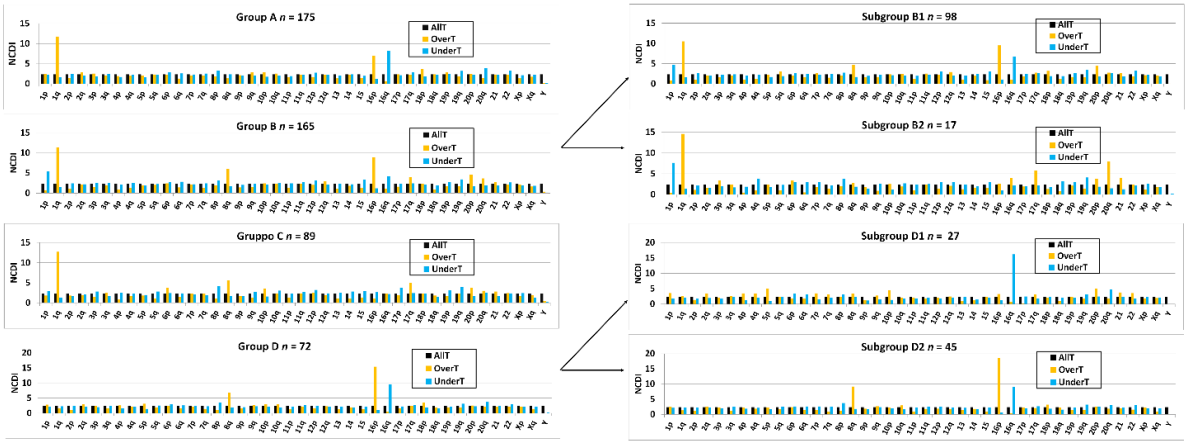


Figure 20. The normalized chromosomal distribution index (NCDI) of OverT (Overexpressed Transcript in comparison to CTRL) and UnderT (Underexpressed Transcript in comparison to CTRL of each 1,16-chromogroup. NCDI values of all transcripts (AllT) analyzed by RNA-seq in each chromosomal arm are also reported for comparison.

In **Table 7**, we report the number of 1q-OverT (OverT encoded by genes located in chromosome 1q; FCvsCTRL > 1.3; adjp < 0.05) and 1q-UnderT (UnderT encoded by genes located in 1q; FCvsCTRL < −1.3; adjp < 0.05) in the examined 1,16-chromogroups. The OverT/UnderT ratio clearly distinguished groups bearing a combined 1q-gain and 16q-loss (A and B1) from those bearing only the 1q-gain (B2 and C) and those showing only a 16q-loss (D1 and D2). The number of 16q-OverT and 16q-UnderT are also reported in the same **Table 7**. As expected, an increased number of 16q-UnderT can be observed in groups bearing the 16q-loss (groups A, B1, D1, and D2).

	1q-OverT*	shared with A	% of A	1q-UnderT**	shared with A	% of A	ratio OverT/ UnderT
A	756	756	100	180	180	100	4.20
B1	737	634	83.86	180	130	72.22	4.09
B2	452	387	51.19	31	22	12.22	14.58
C	727	557	73.68	68	45	25.00	10.69
D1	34	29	3.84	64	50	27.78	0.53
D2	64	42	5.56	204	114	63.33	0.31
	16q-OverT*	**shared with A**	**% of A**	**16q-UnderT****	**shared with A**	**% of A**	**ratio OverT/ UnderT**
A	20	20	2.65	418	418	100	0.05
B1	32	12	1.59	328	281	67.22	0.10
B2	55	7	0.93	22	19	4.55	2.50
C	60	10	1.32	48	40	9.57	1.25
D1	4	2	0.26	216	212	50.72	0.02
D2	10	5	0.66	354	301	72.01	0.03

***FCvsCTRL**>1.3, adjp<0.05; ** **FCvsCTRL**<-1.3, adjp<0.05.

Table 7. Number of 1q- and 16q-OverT and 1q-and 16q-UnderT in 1,16-Chromogroups.

4.1.3 Hierarchical Clustering of Significant OverT and UnderT

We selected DEGs (FCvsCTRL > 1.3 or <−1.3 at adjp < 0.001) between group A and the CTRL group. We focused on DEGs located on chr1 and chr16, thus obtaining a list of 1471 DEGs that were denominated "1,16-A-DEGs" and that comprise 830 OverT (FCvsCTRL > 1.3) and 641 UnderT (FCvsCTRL < −1.3). In order to obtain a global comparison of the expression of 1,16-A-DEGs in 1,16-chromogroups, we performed a hierarchical clustering analysis using the "modified FCvsCTRL values" of the 1471 genes for each group (**Figure 21**; chromogroups in columns). A clear clustering of the groups according to similarity in the chromosome aberration pattern was observed. Group A and B1 clustered together in agreement with the fact that those groups harbor a concomitant 1q gain and 16q loss; group C and B2 clustered in accordance to the shared 1q-gain and disomic chr16, while the clustering

of D1 and D2 reflected the shared 16q-loss and disomic chr1. The clustering at the gene level (rows in **Figure 21**) clearly showed the formation of clusters corresponding to gene expression changes concordant with copy number changes. For instance, gene cluster 4 was found to contain genes located in chromosome 1q and overexpressed in 1q gain-bearing groups, gene cluster 1 was found to mainly be composed of underexpressed genes in 16q-loss samples, and gene cluster 5 was found to contain overexpressed genes in 16p-gain samples. Interestingly, clusters 3 and 2 were characterized by genes, mainly localized in chromosome 1q, that showed a higher expression in groups A and B1, the two groups simultaneously bearing 1q-gain and 16q-loss. The values of the "modified FCvsCTRL" are reported in **Figure 22** for some representative genes belonging to clusters 1, 3, and 4.

4.1.4 Integrated Cytogenomics and Transcriptomic Analysis

The global analysis in the previous paragraphs confirmed that the transcriptional gene-dosage effect is quantitatively relevant in establishing gene expression differences among the various 1,16-chromogroups, but it did not provide any hints on the molecular mechanisms involved in generating putative cancer driver effects. The sensitivity of different genes to the dosage effect may depend on the cellular context (Pollack et al., 2002; Upender et al., 2004) and, in the case of cancer cells, on the specific cancer type or subtype. Moreover, it is likely that only a subgroup of those "dosage sensitive genes," affected by recurrent cancer type-specific aneuploidies, are exerting driver effects. Several strategies have been previously devised to identify such genes (Aure et al., 2013; Davoli et al., 2013, 2017; Hawthorn et al., 2010; Srihari et al., 2016; Taylor et al., 2018).

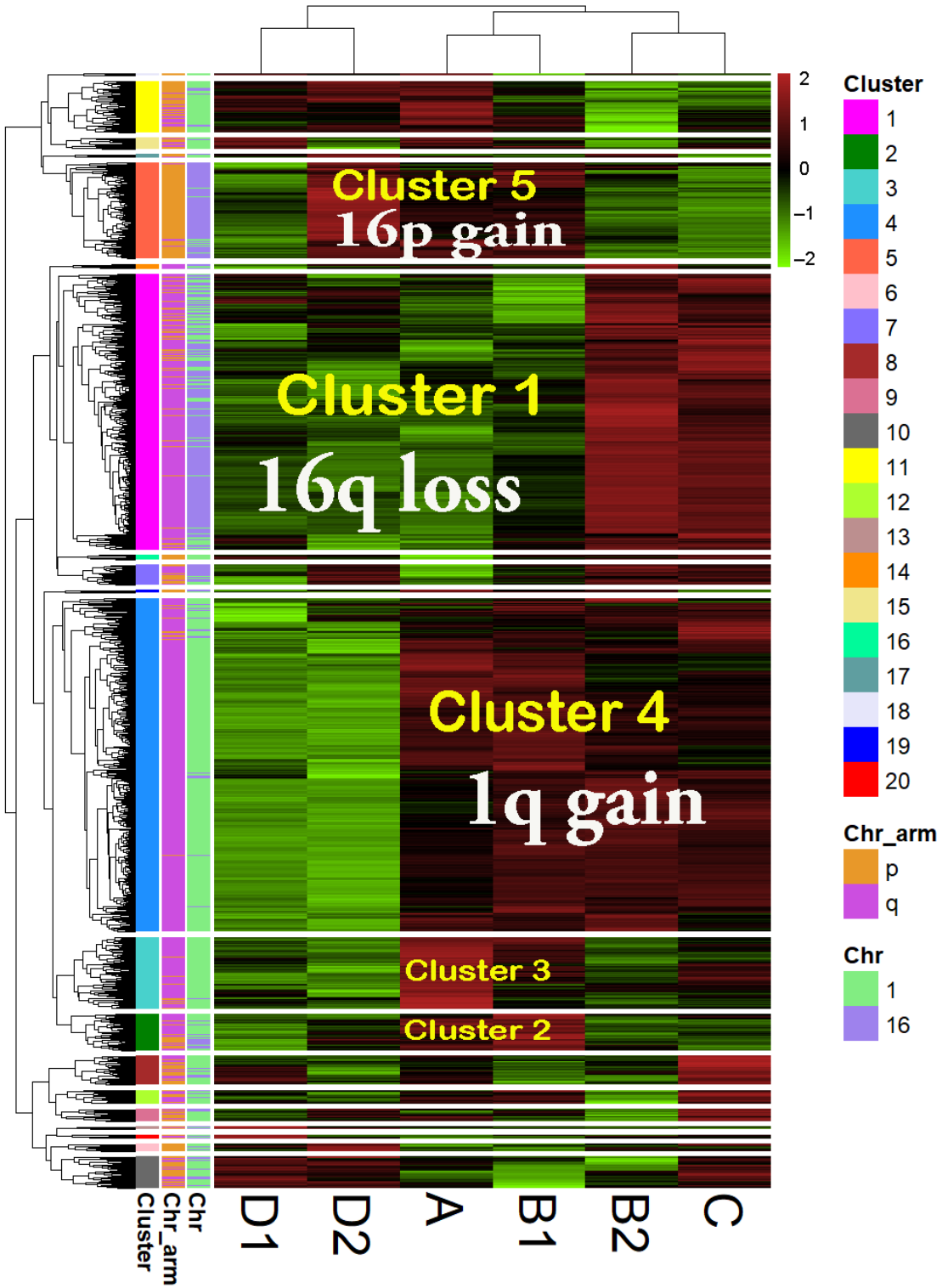


Figure 21. Hierarchical clustering of 1471 differentially expressed genes (DEGs) comparing group A vs. CTRL (FCvsCTRL > 1.3 or <−1.3 at adjp < 0.001) Only DEGs located on chromosome 1 and 16 were used for analysis. The transformation of "linear FC" to "modified linear FC" (modified linear FC is equal to "linear FC-1" if linear FC > 1 or to "linear FC + 1" if linear FC < 1) was performed in order to avoid the gap between −1 and +1 present in linear FC values. Data are clustered by the unweighted pair group method with arithmetic mean (UPGMA) with Euclidean distance. The common chromosomal abnormality among different groups is overwritten on the corresponding columns (in white letter). The cluster number is indicated by colors in the first column on the left and is repeated in some cases by overwriting the data columns (in yellow letters). Chr: chromosome.

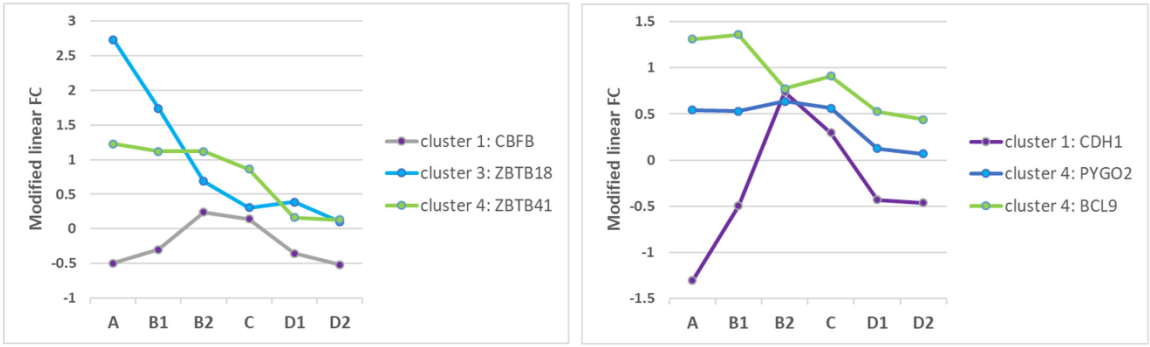


Figure 22 Values of the "modified linear FCvsCTRL" for some representative genes belonging to cluster 1, 3 and 4.

In the present work, our strategy was founded on the hypothesis that functional interactions between the products of genes located in 1q and 16q can underlie the cancer driver effects of co-occurrent 1q-gain and 16q-loss. Such functional interactions can take place at several different levels, such as transcriptional regulation, non-coding RNAs interactions, protein interactions, post-translational modifications, post-translational degradations, and metabolic pathways. Transcriptome data from the different chromogroups and some arguments based on previous studies can provide a guide for the selection of putative driver genes located on 1q or 16q. The first argument is that some cancer driver genes are, indeed, located in chromosome arms 1q and 16q, and data supporting this statement were already reported in the introduction. The second argument is that transcriptional changes driven by gene-dosage are the only effects shown to be induced by der(1;16) or i(1q) up to now; other mechanisms, such as gene fusions, enhancer hijacking, or chromatin transcriptional dysregulation, have not been detected. The third argument is that cancer driver genes located on 1q/16q are likely to be differentially expressed, at the transcript level, between 1,16-chromogroup bearing 1q-gain and 16q-loss and the "CTRL group."

Taking such arguments into account, what are the appropriate descriptors of the transcriptional dysregulation of 1q or 16q genes in the different 1,16-chromogroups? In the previous paragraph, we used the terms "overexpression" and "underexpression" to indicate increases or decreases of transcript levels comparing cancers bearing an arm-level aberration of chromosome 1 and 16 with a so-called CTRL group. The corresponding transcripts have been denominated as "OverT" and UnderT," and the parameter used to describe quantitatively such differences has been called "fold-change vs. control" (FCvsCTRL). However, another important descriptor of cancer dysregulation relies on a comparison between the cancer group and the corresponding normal tissue. Increased or decreased transcripts in such comparison are indicated here as "Upregulated Transcripts" (UpT) or Downregulated Transcripts

(DownT), and the corresponding quantitative parameter is the "fold-change vs. normal tissue" (FCvsN). In the following paragraphs, we use these parameters, alone or in combination, in order to select putative candidate driver genes associated with transcriptional increases linked to 1q-gain or to transcriptional decreases linked to 16q-loss.

As reported in **Table 8**, the three 1,16-chromogroups bearing a diploid 1q (D1, D2, and CTRL) showed an approximately equal number of 1q-located UpT and DownT (UpT/DownT ratio ranging from 1.0 to 1.6), while groups bearing the 1q-gain (A, B1, B2, and C) showed a higher number of UpT (UpT/DownT ratio ranging from 3.01 to 3.45). A relatively large number of genes (about 25–50% of cancer upregulated transcripts detected in 1q) were found to be specifically upregulated in 1q-gain bearing cancers. The number of UpT and DownT in chr16q are also reported in **Table 8**: as expected, the ratio UpT/DownT in 16q was lower in cancer groups bearing the 16q-loss (groups A, B1, D1, and D2).

1,16-Chromogroups	1q-UpT *	Shared with A	% of A	Shared with CTRL	% of CTRL	1q-DownT **	Shared with A	% of A	Shared with CTRL	% of CTRL	UpT/DownT Ratio
A	939	939	100	525	88.53	312	312	100	229	61.73	3.01
B1	897	840	89.46	526	88.70	298	259	83.01	218	58.76	3.01
B2	744	690	73.48	511	86.17	241	204	65.38	196	52.83	3.09
C	898	798	84.98	550	92.75	260	225	72.12	210	56.60	3.45
D1	487	476	50.69	426	71.84	373	261	83.65	277	74.66	1.31
D2	480	469	49.95	423	71.33	479	284	91.03	320	86.25	1.00
CTRL	593	525	55.91	593	100	371	229	73.40	371	100	1.60
	16q-UpT *	**Shared with A**	**% of A**	**Shared with CTRL**	**% of CTRL**	**16q-DownT ****	**Shared with A**	**% of A**	**Shared with CTRL**	**% of CTRL**	**UpT/DownT Ratio**
A	138	138	100	129	40.44	375	375	100	135	92.47	0.37
B1	179	126	91.30	162	50.78	320	300	80.00	135	92.47	0.56
B2	301	118	85.51	251	78.68	117	110	29.33	90	61.64	2.57
C	329	128	92.75	275	86.21	164	161	42.93	122	83.56	2.01
D1	104	89	64.49	102	31.97	290	280	74.67	129	88.36	0.36
D2	129	102	73.91	119	37.30	359	309	82.40	145	99.32	0.36
CTRL	319	129	93.48	319	100	146	135	36.00	146	100	2.18

* FCvsN > 1, adjp < 0.05; ** FCvsN < −1, adjp < 0.05.

Table 8. Number of 1q- and 16q-UpT (Upregulated Transcripts) and 1q and 16q-DownT (Downregulated Transcripts) in 1,16-chromogroups.

Gene expression data were also analyzed by focusing on the comparison between the CTRL cancer group and each of the other 1,16-chromogroups, as described in the previous sections. Such comparison allowed for the classification of transcripts as OverT (overexpressed versus CTRL, FCvsCTRL > 1.3 at adjp < 0.05), a class enriched in gene-dosage-sensitive genes. However, OverT could be upregulated (OverUpT), downregulated (OverDownT), or not significantly changed in comparison to normal tissue. On the basis of previous studies (Condorelli et al., 2018; Condorelli et al., 2019), we only prioritized those genes that are

overexpressed (FCvsCTRL > 1.3, adjp < 0.05) and upregulated (FCvsN > 1, adjp < 0.05) as candidate driver genes located in gained 1q and called them 1q-OverUpT. Out of 2410 transcripts located in chromosome 1q, 639 transcripts could be classified as OverUpT in group A (1q-A-OverUpT). Indeed, after identifying 1q-OverUpT in all chromogroups, we observed that a large number of genes were shared among 1q-gain groups: 437 among groups A, B1, and C and 436 among groups A, B, and C (**Figure 23A**, left panel).

Moreover, the following nineteen 1q-OverUpT genes are also shared with the Group D (not bearing 1q-gain): TRIM46, SLC19A2, BCL9, CRABP2, SOX13, AL136987.1, SLC30A1, CIART, DENND1B, C1orf100, AL391001.1, TUFT1, TRAF5, PIAS3, AGT, IL19, CFAP45, F13B, and ANXA9. Those genes were called "core 1q-OverUpT," and the expression values, reported as TPM, of representative transcripts are shown in **Figure 23B**.

In order to select candidate genes located in 16q that can cooperate with 1q-OverUpT, we prepared a list of UnderT genes in the comparison group A vs. control (FCvsCTRL < −1.3, adjp < 0.05). In the case of 16q-loss, the differential expression against the normal tissue was not included among the selection criteria because, as previously discussed by Condorelli et al. 2018 (Condorelli et al., 2018), a decreased expression of both upregulated and downregulated genes might play a significant role in cancer progression. Out of 1078 transcripts located in chromosome 16q, 418 transcripts could be classified as UnderT in group A (16q-A-UnderT). Indeed, after identifying 16q-UnderT in all 1,16-chromogroups we observed that a large number of genes (n = 208) were shared among 16q-loss groups (A, B1, and D) and other 80 genes among group A and D, as shown in the Venn diagram in **Figure 23A** (right panel).

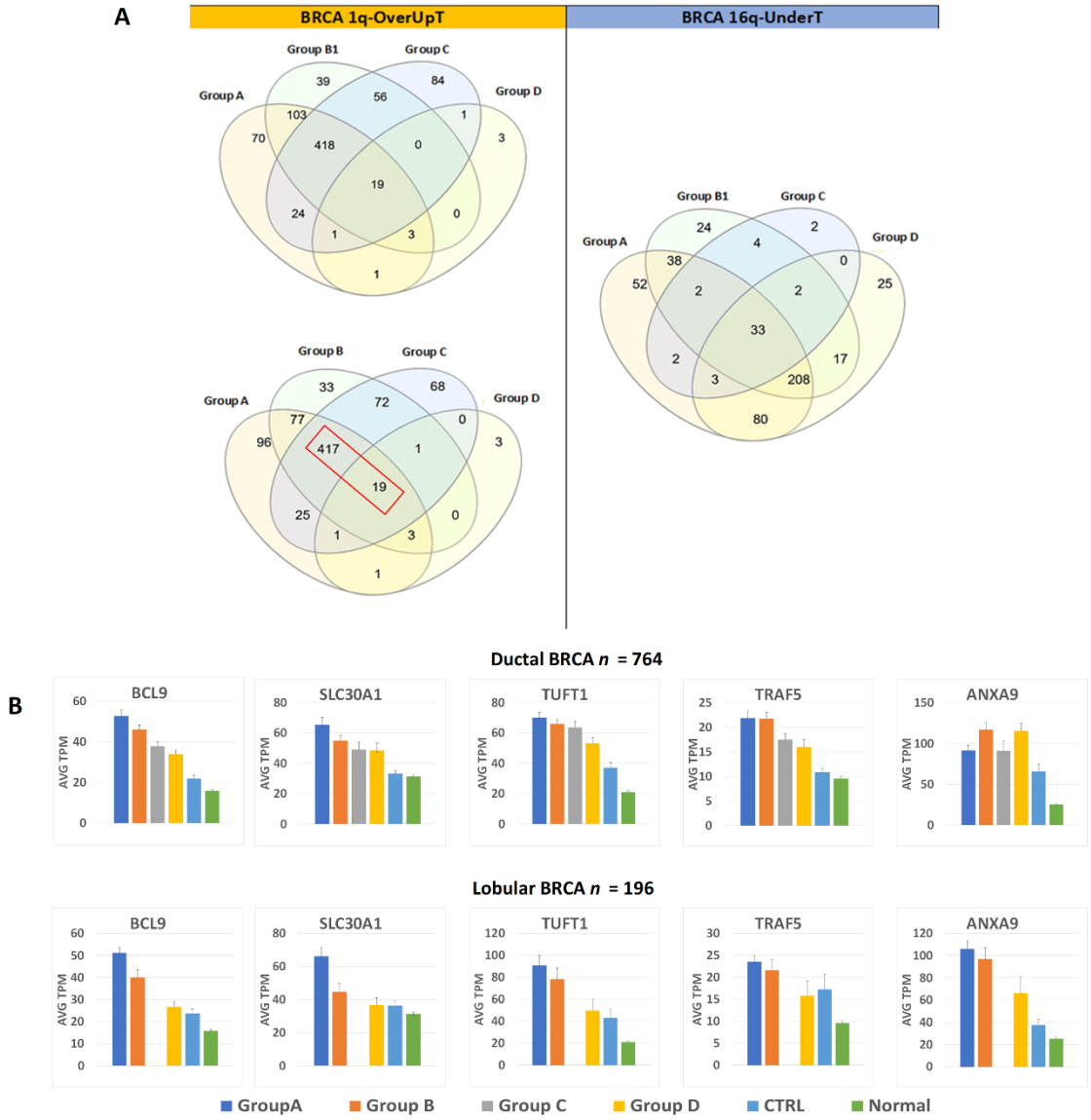


Figure 23. (A) Venn diagrams showing the number of shared 1q-OverUpT (transcripts that are overexpressed in comparison to CTRL and upregulated in comparison to normal tissue) (left panel) and 16q-UnderT (right panel) among different 1,16-chromogroups. The 1q-OverUpT shared among groups A, B, and C are indicated by a red box (436 genes). (B) Representative core 1q-OverUpT in ductal or lobular breast carcinomas; expression values are reported as averages (AVG) of TPM (transcripts per million) ± SEM. The total number of ductal and lobular cancer samples (*n*) is reported in the graph. Due to the rarity of lobular histotype in group C (one sample), no average value could be calculated and the corresponding column is absent in the graph.

4.1.5 Pathway Enrichment Analysis of 1q-OverUpT and 16q-UnderT Genes

In order to obtain information on functional pathways linked to 1q-OverUpT genes in group A, we submitted the list of 639 1q-A-OverUpT genes (FCvsCTRL > 1.3 at adjp < 0.05 and FCvsN > 1 at adjp < 0.05) to Metascape (Zhou et al., 2019) and performed a pathway enrichment analysis and a protein–protein interaction analysis (PPI). Metascape identified all statistically enriched terms in the list using different knowledge-bases and the top 20 pathways are reported in **Figure 24A**. A similar pathway enrichment analysis was performed using the list of 436 1q-OverUpT genes (indicated by a red box in **Figure 23A**) shared by all 1q-gain groups A, B, and C. As shown in **Figure 24B**, most of the top-ranking genes were shared among the two analysis, with the APH1–PSEN2–NCSTN complex as the top-first pathways in both lists.

In order to explore the functional cooperation among 1q and 16q genes, a combined list of 639 1q-A-OverUpT genes (FCvsCTRL > 1.3 at adjp < 0.05 and FCvsN >1 at adjp < 0.05) and 418 16q-A-UnderT genes was submitted to Metascape. The results of the pathway enrichment analysis are shown in **Figure 24C** (top 20 pathways). An analysis of pathways including 1q and 16q genes revealed a cooperation of the WWP2 gene (chr16q) in the "APH1A–PSEN2–NCSTN complex" (CORUM: 2735; one of the top-20 pathways in **Figure 24**) and "NOTCH3 Activation and Transmission of Signal to the Nucleus" (R-HSA-9013507) pathways.

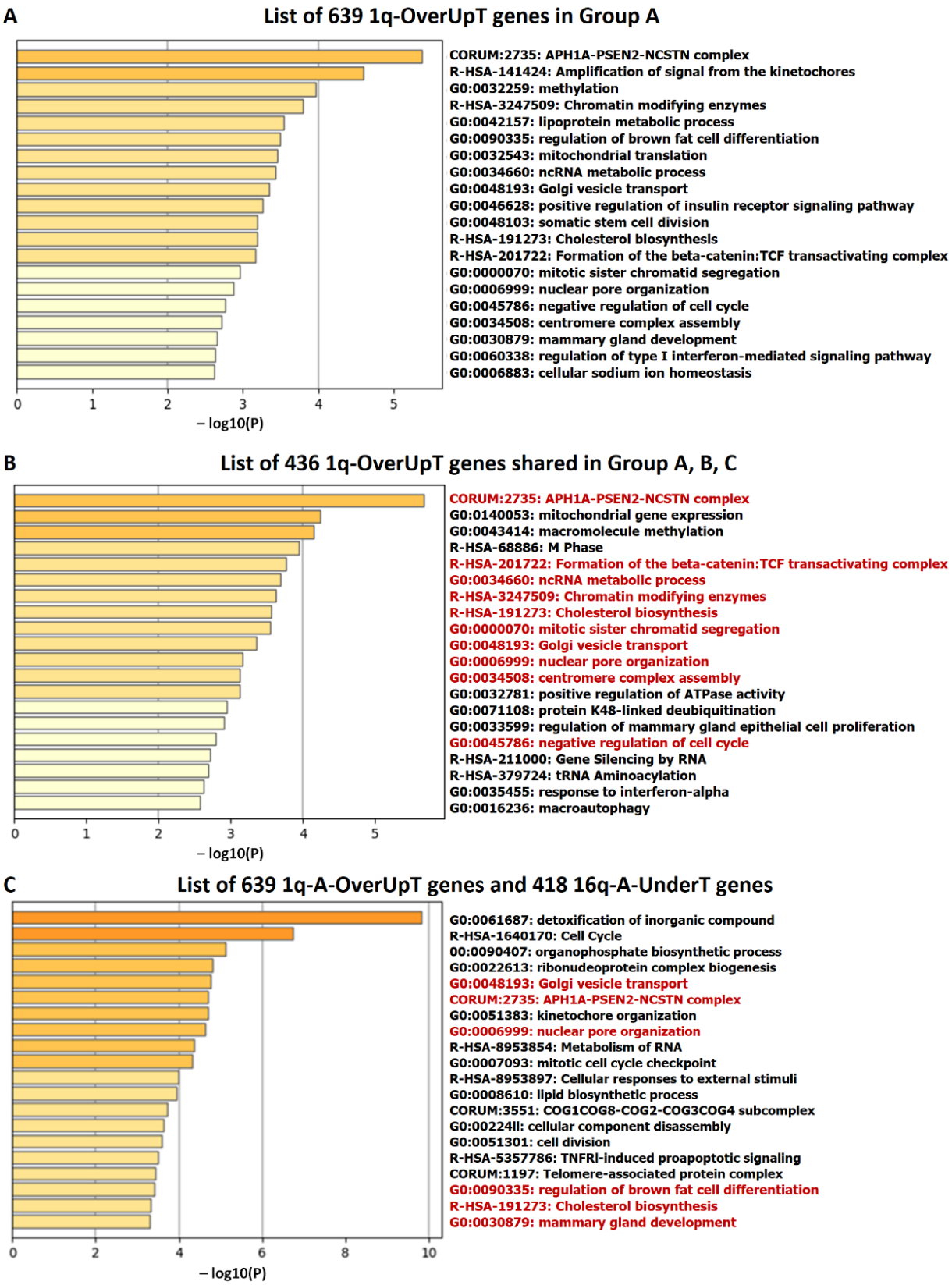


Figure 23 Pathway enrichment analysis using: (A) the list of 639 1q-OverUpT genes in group A, (B) the list of 436 1q-OverUpT genes shared among the three 1q-gain groups, and (C) the combined list of 639 1q-A-OverUpT genes (FCvsCTRL > 1.3 at adjp < 0.05 and FCvsN >1 at adjp < 0.05) and 418 16q-A-UnderT genes. Pathways in red letters in (B,C) are present also in (A).

4.1.6 Gene Set Enrichment Analysis (GSEA)

In a previous subsection, we determined whether the list of a special group of differentially expressed genes, called 1q-OverUpT and 16q-UnderT, was enriched for pathway or ontology terms. As is the case with over-representation methods, the results were dependent on the cutoff used in constructing the list. In this subsection, we report the results obtained with an additional tool for the analysis of genome-wide gene expression profiles, called GSEA (Subramanian et al., 2005). GSEA takes the expression values of all transcripts, not only those above an arbitrary cutoff of fold-change or significance, into account. TPM values obtained in RNA seq analysis of 56,000 transcripts from 1058 samples were given as input to the GSEA software, and each cytogenetic group was compared with the CTRL group (Analysis I). Comparing group A vs. CTRL, 21 pathways from the REACTOME database showed an NES >1.5 (**Table 9**), although none of those pathway showed an FDR (False Discovery Rate) value <0.25. As indicated in **Table 9**, 20 out of 21 top-pathways included genes located in chromosome 1q among the leading edge subset.

Pathway Name	Size	ES	NES
* NRIF_SIGNALS_CELL_DEATH_FROM_THE_NUCLEUS	16	0.710	1.828
* NOTCH2_ACTIVATION_AND_TRANSMISSION_OF_SIGNAL_TO_THE_NUCLEUS	22	0.620	1.767
*° MITOCHONDRIAL_TRNA_AMINOACYLATION	18	0.776	1.754
*° SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	18	0.736	1.746
*° MITOCHONDRIAL_FATTY_ACID_BETA_OXIDATION	36	0.556	1.634
*° TP53_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_CYTOCHROME_C_RELEASE	20	0.597	1.626
*° ENERGY_DEPENDENT_REGULATION_OF_MTOR_BY_LKB1_AMPK	29	0.588	1.610
* DEACTIVATION_OF_THE_BETA_CATENIN_TRANSACTIVATING_COMPLEX	42	0.552	1.608
* ZINC_TRANSPORTERS	17	0.582	1.599
*° FORMATION_OF_INCISION_COMPLEX_IN_GG_NER	43	0.595	1.587
*° INTRAFLAGELLAR_TRANSPORT	54	0.553	1.580
*° TRNA_AMINOACYLATION	24	0.667	1.576
*° SUMOYLATION_OF_DNA_METHYLATION_PROTEINS	16	0.663	1.569
*° CILIUM_ASSEMBLY	200	0.528	1.568
*° PEROXISOMAL_PROTEIN_IMPORT	63	0.518	1.543
*° ANCHORING_OF_THE_BASAL_BODY_TO_THE_PLASMA_MEMBRANE	96	0.545	1.542
° DISEASES_ASSOCIATED_WITH_N_GLYCOSYLATION_OF_PROTEINS	17	0.628	1.541
*° RESOLUTION_OF_ABASIC_SITES_AP_SITES	38	0.608	1.535
*° RAB_GERANYLGERANYLATION	65	0.488	1.528
*° RNA_POLYMERASE_III_TRANSCRIPTION_TERMINATION	23	0.589	1.528
* CELL_DEATH_SIGNALLING_VIA_NRAGE_NRIF_AND_NADE	76	0.497	1.528

* indicates pathways including 1q genes among the leading edge genes; ° indicates pathways including 16p genes among the leading edge genes.

Table 9. GSEA (gene set enrichment analysis) analysis of cytogenetic group A vs. CTRL (1058 samples; Analysis I). Reactome pathways with normalized enrichment score (NES) > 1.5.

Clues about functional interactions among transcriptionally dysregulated genes could be derived by the analysis of the pathways "NOTCH2 Activation and Transmission of Signal to the Nucleus," "Deactivation of the beta-catenin transactivating complex," and "Formation of the beta-catenin:TCF transactivating complex" (**Table 9**). Interestingly, the top-ranked genes

in the two beta-catenin pathways were BCL9, PYGO2, RBBP5, and CDC73. The top-ranked genes in the "NOTCH2 Activation and Transmission of Signal to the Nucleus" were the genes APH1A, PSEN2, and NCSTN, whose products were subunits of the γ-secretase, a protease complex able to cleave various proteins within their transmembrane domains. Indeed, an increased expression of those genes in hormone receptor-positive breast cancers has been previously reported (Filipović et al., 2011; Peltonen et al., 2013). All these genes belong to the 1q-OverUpT group, and their pathways were already identified in the Metascape analysis reported in the previous subsection.

Among shared pathways between groups A and D or among groups A, B1, and D, several pathways, such as "Mitochondrial tRNA aminoacylation" and "Formation of Incision Complex in GG-NER," include 16p- and 1q-located genes among the leading edge subset of GSEA analysis of both group A and D, and several genes, such as PARP1, IARS2, TARS2, and DARS2 belong to the 1q-OverUpT group. This result further confirmed that some 1q genes involved in cancer-activated pathways were also overexpressed in a cytogenetic group devoid of 1q-gain (group D) and might functionally cooperate with genes located in 16p. Therefore, such pathways may provide some clues about functional interactions induced by 1q-gain and 16p-gain aberrations. Indeed, 16p gain was found to be a frequent aberration shared by tumours of groups A, B1, and D. However, such GSEA analysis could not provide information about functional interactions between 1q and 16q genes because genes were ranked according to the real value of Signal2Noise (i.e., the difference of means of the two compared groups scaled by the standard deviation), a metric score that could take either positive or negative values. Genes modified by the transcriptional dosage-effect were expected to show positive scores if located in 1q-gain and negative scores if located in 16q-loss and to be ranked at the opposite ends of the list, thus preventing the identification of putative cooperative effect of 1q-OverT and 16q-UnderT genes. In order to overcome this issue, we repeated the analysis by sorting the genes using the absolute value of the Signal2Noise metric score. Many of the functional pathways associated with cytogenetic groups in the previous analysis (real value-analysis) were confirmed in the analysis based on the absolute value of Signal2Noise. In the group A analysis, "Formation of the beta-catenin:TCF transactivating complex," "Deactivation of the beta-catenin transactivating complex," "Mitochondrial tRNA aminoacylation," "NOTCH2 Activation and Transmission of Signal to the Nucleus," "NOTCH3 Activation and Transmission of Signal to the Nucleus," and "Signaling by NOTCH2" showed an NES > 1.5. In this analysis, the pathway "Ephrin mediated repulsion of cells" ranked among the top positions due to the 1q-located genes

EFNA1, EFNA4, and EFNA3 and the gamma secretase components APH1A, PSEN2, and NCSTN. Indeed, several pathways were shared among different cytogenetics groups The pathway "NOTCH3 Activation and Transmission of Signal to the Nucleus" was enriched in group A (NES = 1.71, nominal *p*-value = 0.001, FDR *q*-value = 0.28), B1 (NES = 1.75, nominal *p*-value < 0.001, FDR *q*-value = 0.11), group B2 (NES = 1.52, nominal *p*-value = 0.005, FDR *q*-value = 0.30), group C (NES = 1.75, nominal *p*-value < 0.001, FDR *q*-value = 0.04), and group D2 (NES = 1.61, nominal *p*-value < 0.001, FDR *q*-value = 0.12).

The absolute score-analysis allowed for the identification of pathways including 1q and 16q-located genes among the leading-edge subsets. After excluding the pathways showing a discordant functional effect of 1q-OverT genes and 16q-UnderT genes, such analysis suggested possible cooperative functional interactions in the pathway "NOTCH3 Activation and Transmission of Signal to the Nucleus" in groups A, B1, and D2. In the "NOTCH3 Activation and Transmission of Signal to the Nucleus" pathway, the overexpression of three 1q genes encoding subunits of the gamma secretase complex (APH1A, PSEN2, and NCSTN), which plays a positive role in NOTCH3 signalling, was found to be functionally interconnected with the reduced expression of the 16q gene WWP2, an E3 ubiquitin ligase that negatively regulates NOTCH3 signalling (Jung et al., 2014). Another pathway implicating a functional cooperation between 1q genes and 16q genes was the "Nuclear Signaling by ERBB4" enriched in group D2 (NES = 1.61, nominal *p*-value: <0.001, FDR *q*-value = 0.11). The involved 1q gene was APH1A and the 16q gene was WWOX, a WW-domain-containing protein that binds to a cytosolic fragment of ERBB4 (generated by the gamma secretase complex) and prevents its translocation to the nucleus (Aqeilan et al., 2005).

4.1.7. Analysis II. 1,16 Chromogroups: Ductal LumA Adenocarcinomas

In this work, we compared tumours bearing a specific arm-level aberration (study group) to tumours not bearing it (control group). The basic assumption was that part of the gene expression differences between the two groups were linked to the presence of the chromosomal aberration. However, the choice of criteria used for the generation of the control cancer group exerted a strong impact on the results. Indeed, two important points should be taken into account: (1) although tumours were found to belong to the same clinico-pathological type (breast invasive carcinoma), phenotypic and genotypic heterogeneity was present in all cytogenetic groups; (2) cancers in the so-called "control group" differed from those in the study group, not only for the absence of the specific chromosomal aberration but also for other mutational events and carcinogenesis pathways. Regarding the first point, it is

possible to classify breast invasive carcinomas in histological subtypes using classical optical microscopy or in molecular subtypes using specific molecular biomarkers and transcriptome analysis (Ciriello et al., 2015; Koboldt et al., 2012). The frequencies of histological subtypes and molecular subtypes in the different cytogenetic groups are reported in **Figure 25**. Invasive ductal carcinoma was found to be the predominant histological subtype in all cytogenetic groups. Invasive lobular carcinomas showed the highest frequency in group A and were associated with cytogenetic groups bearing 16q-loss (A, B1, D1, and D2). Molecular subtype LumA was enriched in cytogenetic groups A, B1, D1, and D2, while LumB was enriched in subgroup B2. The basal-like subtype showed the highest frequency in the cytogenetic group C, while the normal-like subtype showed the highest frequency in the control group.

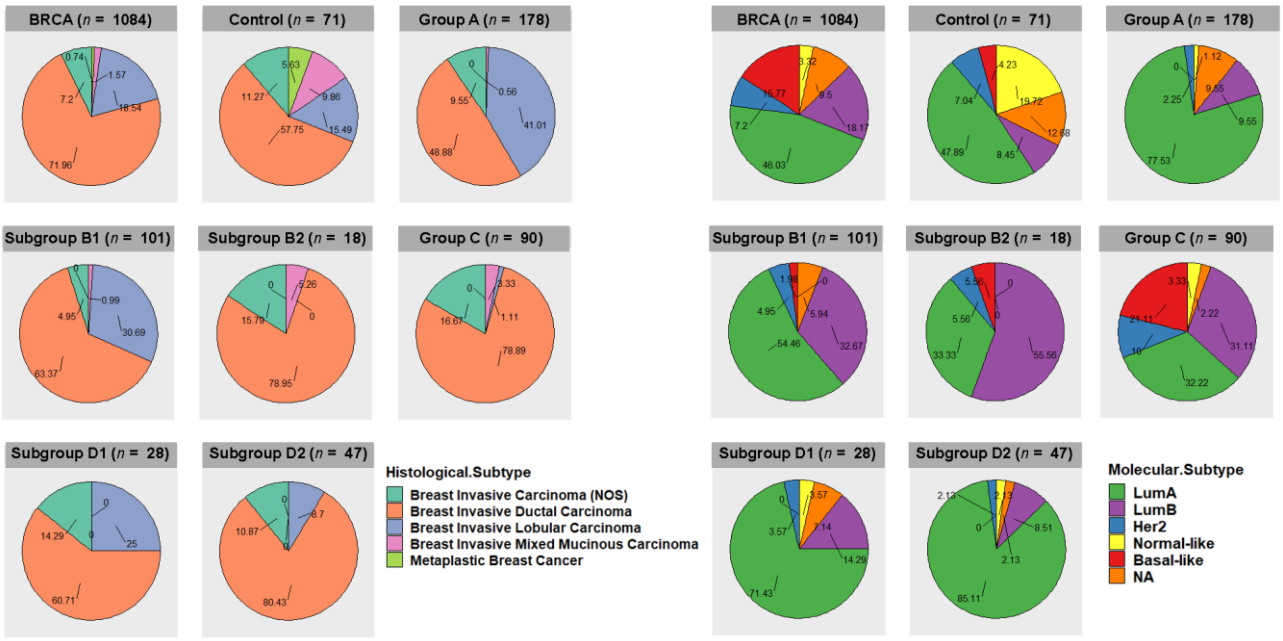


Figure 25. Frequencies of histological subtypes (left panels) and molecular subtypes (right panels) in the different 1,16-chromogroups.

Therefore, in order to reduce phenotypic and genotypic heterogeneity, we repeated all the transcriptome analyses by selecting only the most common histological and molecular subtypes (breast ductal carcinoma and LumA, respectively). The available number of breast ductal LumA carcinomas allowed for the generation of group A, B, B1, C, D, and D2 (**Figure 26**). As expected, the NCDI values of 1q and 16p were increased in samples bearing 1q-gain and 16p-gain. The increased NCDI value of 8q was still detectable in subgroup D2.

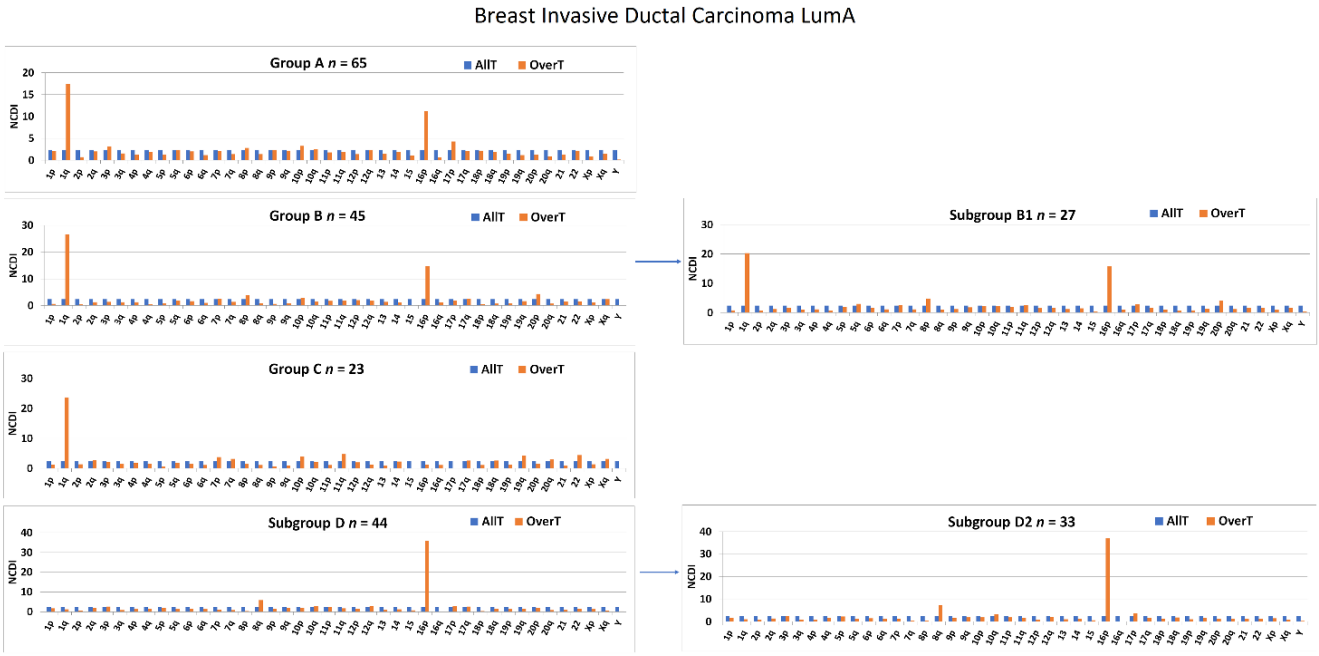


Figure 26. NCDI values of OverT of each 1,16-chromogroup formed with breast invasive ductal carcinomas of the LumA subtype. NCDI values of all transcripts (AllT) analyzed by RNA-seq in each chromosomal arm are also reported for comparison.

The GSEA analysis was repeated using transcriptome data derived from 310 breast ductal LumA carcinomas (herein called "Analysis II" in order to distinguish it from "Analysis I" reported in the previous subsection. In Analysis II, the comparison of group A vs. CTRL identified 23 REACTOME pathways showing an NES >1.48 (**Table 10**).

Pathway Name	Size	ES	NES
NRIF_SIGNALS_CELL_DEATH_FROM_THE_NUCLEUS	16	0.70	1.74
TP53_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_CYTOCHROME_C_RELEASE	20	0.65	1.68
MITOCHONDRIAL_TRNA_AMINOACYLATION	18	0.77	1.67
DEFECTIVE_C1GALT1C1_CAUSES_TN_POLYAGGLUTINATION_SYNDROME_TNPS	17	0.69	1.66
DEACTIVATION_OF_THE_BETA_CATENIN_TRANSACTIVATING_COMPLEX	42	0.59	1.65
DISEASES_ASSOCIATED_WITH_N_GLYCOSYLATION_OF_PROTEINS	17	0.73	1.64
RNA_POLYMERASE_III_TRANSCRIPTION_TERMINATION	23	0.64	1.61
SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	18	0.71	1.60
RNA_POLYMERASE_III_CHAIN_ELONGATION	18	0.67	1.57
NONSENSE_MEDIATED_DECAY_NMD	116	0.66	1.57
FORMATION_OF_THE_BETA_CATENIN_TCF_TRANSACTIVATING_COMPLEX	31	0.61	1.57
DEFECTIVE_GALNT3_CAUSES_FAMILIAL_HYPERPHOSPHATEMIC_TUMORAL_CALCINOSIS_HFTC	16	0.67	1.56
SELENOAMINO_ACID_METABOLISM	109	0.67	1.56
RESOLUTION_OF_ABASIC_SITES_AP_SITES	38	0.64	1.56
ENERGY_DEPENDENT_REGULATION_OF_MTOR_BY_LKB1_AMPK	29	0.60	1.55
GAP_FILLING_DNA_REPAIR_SYNTHESIS_AND_LIGATION_IN_GG_NER	25	0.63	1.51
SULFUR_AMINO_ACID_METABOLISM	28	0.54	1.51
EUKARYOTIC_TRANSLATION_ELONGATION	94	0.73	1.50
MITOCHONDRIAL_FATTY_ACID_BETA_OXIDATION	36	0.54	1.50
TRNA_AMINOACYLATION	24	0.66	1.49
PROLACTIN_RECEPTOR_SIGNALING	15	0.58	1.49
RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY	102	0.67	1.49
NOTCH2_ACTIVATION_AND_TRANSMISSION_OF_SIGNAL_TO_THE_NUCLEUS	22	0.56	1.49

Table 10. REACTOME pathways showing a normalized enrichment score (NES) >1.48 in GSEA Analysis II.

Analyses I and II showed a good overlapping of the 50 top-NES-ranked pathways in all groups (50% group A; 38% group B1, 42% group D2), except for group C (8%). The inclusion of a relatively large group of basal-like cancers in group C in "Analysis I" was the likely explanation for this discrepancy.

Interestingly, "Analysis II" showed a large number of shared pathways (n = 11) among groups A, B1, and D or A, C, and D (**Figure 27**). Indeed, Analysis II (restricted to Ductal LumA cancers) confirmed several pathways previously identified in "Analysis I," such as "Mitochondrial tRNA aminoacylation," "Deactivation of the beta-catenin transactivating complex," "Formation of the beta-catenin:TCF transactivating complex." Moreover, the enrichment of "Formation of the beta-catenin:TCF transactivating complex" (NES = 1.52, nominal p-value = 0.007; FDR q-value = 0.206) and "EPH-ephrin mediated repulsion of cells" (NES = 1.52, nominal p-value < 0.001; FDR q-value = 0.211) were statistically significant in Analysis II performed with absolute Signal2Noise values.

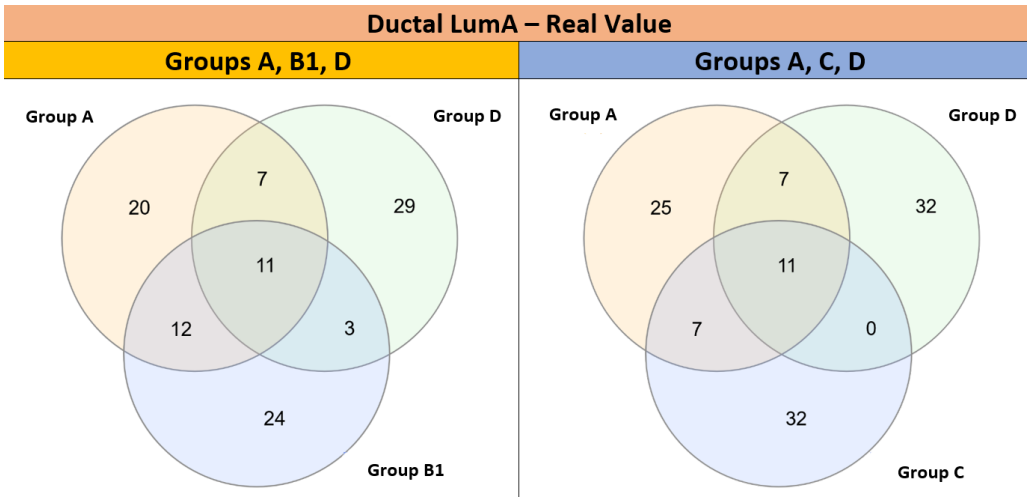


Figure 27. Venn diagrams showing overlapping of top-ranked functional pathways identified in GSEA Analysis II (restricted to 310 invasive ductal carcinomas of the LumA subtype) among different chromogroups.

4.1.8. Analysis III: 1,16-Chromogroups Ductal and Lobular Adenocarcinomas

A detailed molecular characterization of the differences between invasive ductal and lobular carcinomas had been previously provided by integrated omics analysis (Ciriello et al., 2015; Desmedt et al., 2016). The higher frequency of CDH1 loss-of-function mutations and the lower transcriptional expression of the CHD1 gene in lobular carcinomas were some of the main differences between the two histotypes. Indeed, a decreased expression of CDH1, at both the protein and transcript levels, was probably underlying the discohesive phenotype of lobular carcinomas. We confirmed the clear-cut difference in CDH1 expression between ductal and lobular carcinomas (**Figure 28A**). In this work, we could also compare the transcript levels of CDH1 in breast cancer groups differentiated by the presence or absence of Chr1 and 16 aberrations (**Figure 28C**). As already shown in **Figure 25**, the 1,16-chromogroups characterized by 1q-gain and 16q-disomy (groups B2 and C) had no or few invasive lobular carcinomas, thus precluding the analysis of such a cancer histotype in those two groups. However, the CTRL group of ductal carcinomas (i.e., without abnormalities of Chr 1 and 16) showed a sufficient number of both histotypes (40 ductal vs. 11 lobular carcinomas), thus allowing for comparisons with all other breast cancer groups and with corresponding normal tissue. Interestingly, invasive ductal carcinomas with 16q-disomy (CTRL, B2, and C) showed a higher level of CDH1 transcripts in comparison to normal breast tissues, while CDH1 levels in invasive ductal carcinomas with 16q-loss (group A, B1, and D) were similar to those of normal tissue. This observation was in agreement with

previous immunohistochemistry studies showing that low-grade invasive ductal carcinomas have stronger E-cadherin membrane staining than that seen in the normal breast epithelial cells, while E-cadherin loss may occur as a late event in a subgroup of high-grade invasive ductal carcinomas (Alsaleem et al., 2019; Jeschke et al., 2007). On the contrary, E-cadherin loss is observed as an early event in lobular carcinomas (McCart Reed AE, Kutasovic JR, Lakhani SR, 2015). In the present work, we found that CDH1 transcript levels were lower in invasive lobular carcinomas either in 16q-loss groups (A, B1, and D) and in the 16q-disomic CTRL-group when compared to corresponding ductal carcinomas or to normal breast tissue (**Figure 28C**). In other words, invasive lobular carcinomas had a CDH1 expression lower than ductal carcinomas both in the presence (groups A, B1, and D) and the absence of 16q-loss (CTRL group). This observation suggested that additional mechanisms, besides the 16q-loss, are downregulating CDH1 expression in lobular carcinomas. Though it was clear that loss-of-function point mutations of CDH1 are cooperating to the decreased functionality of E-cadherin in lobular carcinomas, the transcript levels of lobular CDH1-mutated cancers were not significantly different from those of CDH1-wild type ones (**Figure 28B**). Indeed, it has been repeatedly suggested that other mechanisms, such as epigenetic modifications, the upregulation of CDH1 transcriptional repressors, and other forms of transcriptional dysregulation, may account for the downregulation of CDH1 transcription in lobular carcinomas (Alsaleem et al., 2019; Ciriello et al., 2015). Our analysis confirmed that the transcriptional downregulation of CDH1 in ductal carcinomas is weaker than that in lobular ones. Though 16q-loss is frequently observed both in ductal and lobular carcinomas (with a slight higher frequency in lobular ones; see **Figure 30**), the CDH1 transcriptional difference between the two histotypes was detectable both in the presence (group A, B1, and D) or the absence of 16q-loss aberrations (CTRL group; **Figure 28C).** Indeed, in the CTRL breast cancer group, it was possible to observe a relevant number of invasive lobular carcinomas in the absence of 16-q loss, and it is interesting to note that this 16q-disomic lobular subtype was found to be characterized by a near-euploid karyotype (**Figure 30**). Nevertheless, 16q-loss is a strong determining factor for the generation of invasive breast lobular carcinomas, as suggested by the lack or the rarity of this histotype in groups B2 and C, bearing 1q-gain but not 16q-loss (**Figure 25**). In conclusion, our analysis suggested that 16-q loss can be considered a critical chromosomal abnormality for the generation of lobular carcinomas in the context of a significant aneuploidy score (>4) (see **Figure 29**, which shows a comparison of arm-level chromosomal aberrations in lobular carcinomas of the CTRL group with those in lobular carcinomas of group A, B1 and D).

Moreover, invasive lobular carcinomas were found to be able to bear 16q-loss either with (group A and B1) or without 1q-gain (group D), in agreement with previous reports (Ciriello et al., 2015; Desmedt et al., 2016). Indeed, the increased frequency of 16q-loss in lobular carcinomas was found to be accompanied by an increased frequency of 1q-gain, confirming that the co-occurrence of 1q-gain and 16q-loss is a feature of both ductal and lobular carcinomas in a larger sample population (**Figure 29**; also note the increased frequency of 8p-loss and 8q-gain in ductal carcinomas).

Statistically significant DEGs between lobular and ductal carcinomas (LvsD-DEGs) were also analyzed by the EdgeR software separately for groups A, B1, and D (decreased in lobular carcinomas: linear fold change lobular vs. ductal <-1.5 and adjp < 0.05; increased in lobular carcinomas: linear fold change lobular vs. ductal > 1.5 and adjp < 0.05). 16q-DEGs shared across all 16q-loss groups (A, B1, and D) are reported in Table S7. As expected (see also **Figure 28C**), CDH1 was significantly decreased in lobular carcinomas vs. ductal carcinomas in all 16q-loss groups (A, B1, and D) and was the only 16q-LvsD-DEGs that was coherently decreased in lobular carcinomas of those three groups. Moreover, CDH1 was the only gene belonging to the class of 16q-UnderT that was further decreased in lobular carcinomas in 16q-loss groups. 16q-LvsD-DEGs were found to represent only a minor fraction of 16q-UnderT: only 3 transcripts out of 208 were LvsD-DEGs (AC040162.3, CDH1, IL34) UnderT shared across groups A, B1, and D (Table S7). In summary the analysis of LvsD-DEGs revealed very few specific transcriptional dysregulations superimposed to the common 16q-loss dependent downregulation, besides the known CDH1 downregulation.

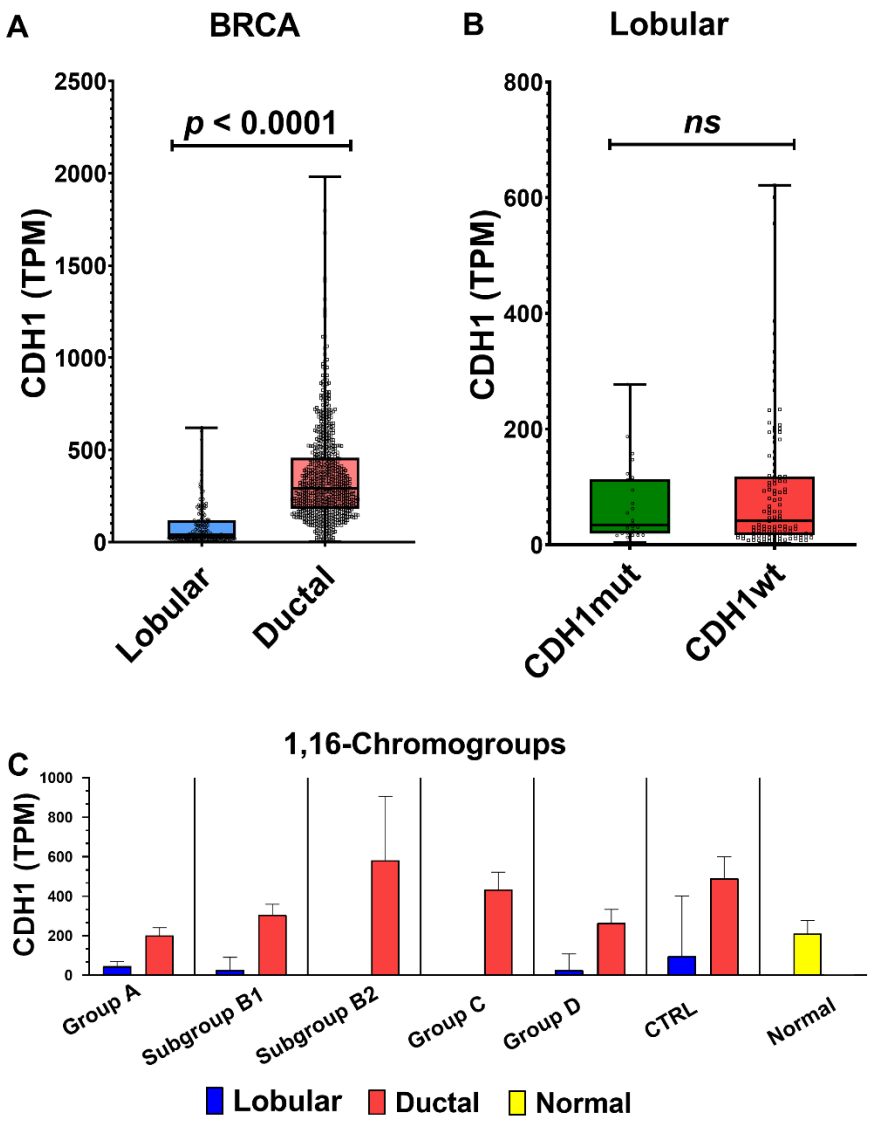


Figure 28. (A) Dot plots of CDH1 transcript levels in lobular and ductal carcinomas and (B) in lobular cancer bearing a mutated CDH1 (CDH1mut) or a wild-type CDH1 (CDH1wt); overlaid boxes show median and interquartile range; statistical significance by Mann–Whitney test; (C) TPM levels of CDH1 in different 1,16-chromogroups differentiated in lobular and ductal cancers. Columns represent the median values, and bars represent the 95% confidence intervals. ns: not statistically significant.

A similar analysis was performed for 1q-LvsD-DEGs by selecting those genes showing concordant changes across 1q-gain groups (A and B1). Again, those 1q-LvsD-DEGs represented only a minor fraction of OverUpT genes shared among groups A and B1 (10 1q-LvsD-DEGs out of 540 OverUpT shared between groups A and B1). Moreover, none of the "core 1q-OverUpT" genes showed a differential expression between ductal and lobular carcinomas, suggesting that their putative functional role in carcinogenesis might be shared between the two histotypes.

Given that the largest number of lobular carcinomas was observed in group A (bearing both 1q-gain and 16q-loss), it is reasonable to hypothesize that cooperative functional networks of 1q and 16q genes could operate both in lobular and ductal carcinomas, the two histotypes

being mainly differentiated by the deeper transcriptional downregulation of 16q-CDH1 in lobular carcinomas.

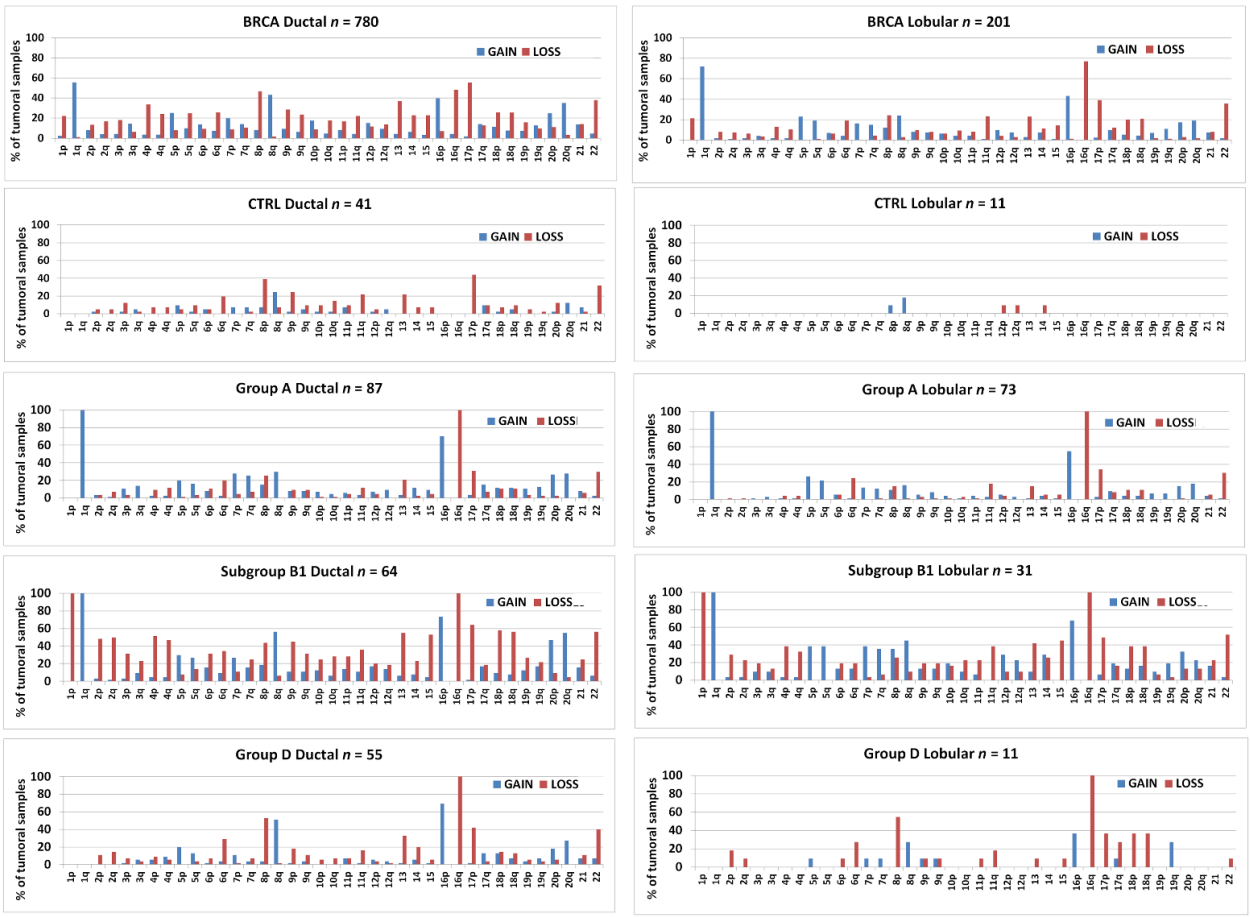


Figure 29. Bar graphs showing the percentage of samples bearing arm-level gains or losses in chromosomes 1–22 in invasive ductal or lobular breast carcinomas or in the different 1,16-chromogroups formed with invasive ductal carcinomas or invasive lobular carcinomas, as indicated in each graphs. The number of samples (*n*) in each chromogroup is reported in the corresponding graph.

4.1.9. Recurrent Point Mutations in Breast Cancer 1,16-Chromogroups

Recurrent point mutations by WES data are shown in oncoplots of **Figures 30 and 31**. The main histological subtypes (ductal or lobular carcinoma) are indicated by the annotation bar below the graph. The higher frequency of TP53 mutations in ductal carcinomas and CDH1 mutations (mainly nonsense or splice-site mutations) in lobular carcinomas was found to be a general feature of those histotypes, as shown in the oncoplot of **Figure 30**, including 709 ductal and 149 lobular BRCA samples analysed by WES in TCGA study. An accurate analysis of recurrent point mutations in invasive ductal and lobular carcinomas has been already provided by previous studies using both TCGA and other data, and it is not repeated here (Banerji et al., 2012; Ciriello et al., 2015; Desmedt et al., 2016).

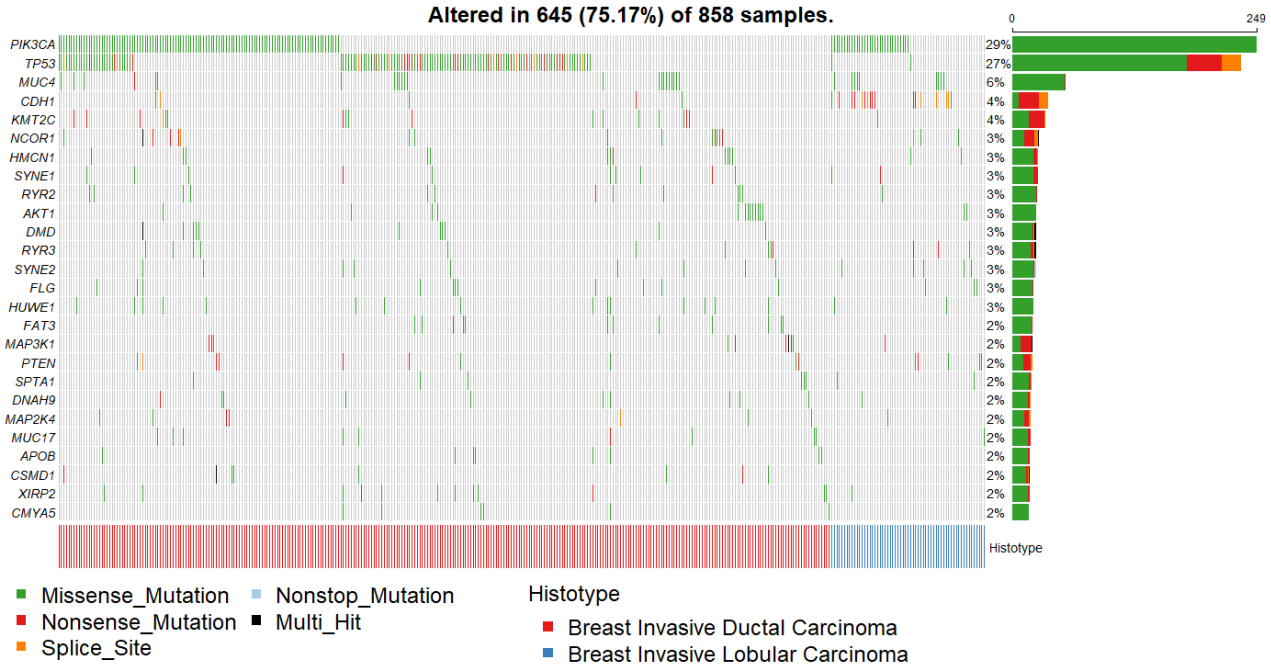


Figure 30 Oncoplot showing point mutations detected by WES in 645 samples out of 709 ductal and 149 lobular BRCA samples analysed in TCGA study.

However, oncoplots in **Figure 31** provide a rapid overview of the top 30 recurrent point mutations detected by WES in the different 1,16-chromogroups examined in the present study. It is clear that this type of presentation did not allow for a direct comparison between ductal and lobular cancers because of the largely different number of samples of the two histotypes in the different chromogroups (note the absence of lobular cancers in Group B2 and C, as already shown **Figure 25**). Nevertheless, a higher frequency of TP53 (nonsense, missense, and splice site) and GATA3 mutations (nonsense and splice site) in ductal carcinomas and CDH1 mutations in lobular carcinomas was easily recognizable in

cytogenetic group A. Mutations of CBFB (mainly missense mutations) were most frequently detected in ductal carcinomas of group A and B1, while mutations of MAP3K1 (nonsense and missense) and ARID1A (missense and splice site mutations) were most frequently detected in ductal carcinomas of group D. PI3K mutations (missense) were found to be the most frequent mutation in all the chromogroups, with the exception of group C, where TP53 mutations (nonsense, splice sites, and missense) predominated in agreement with the higher level of chromosomal aberrations detected in this group. All those mutations were reported as significant in ductal or lobular cancers, or both, in previous analysis of TCGA data (Ciriello et al., 2015).

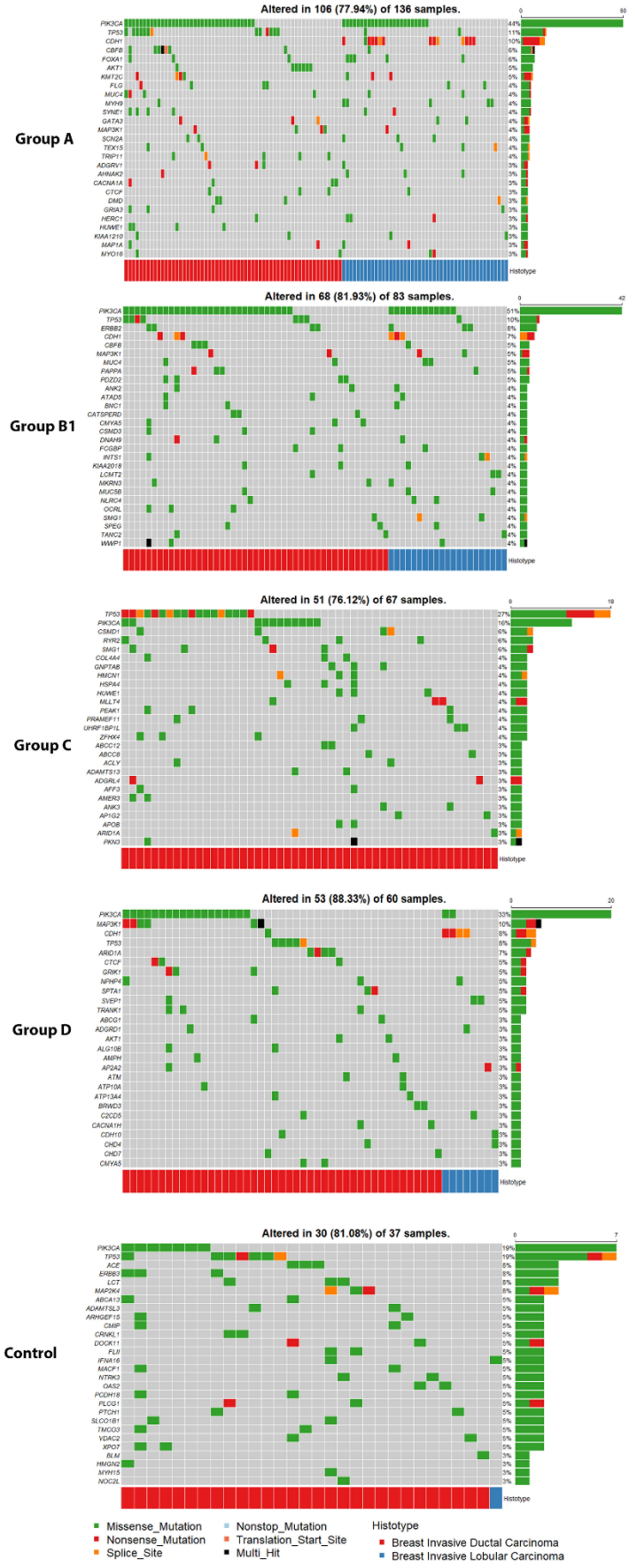


Figure 31. Oncoplots showing point mutations detected by WES in 1,16-chromogroups.

4.1.10 CNA analysis of METABRIC cohort

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study was explored to evaluate the clinical prognosis of breast cancer. Copy number alteration (CNA) data were analysed by Gistic2 (Mermel et al., 2011) (Beroukhim et al., 2010) (Beroukhim et al., 2007). The algorithm assigns a numeric value (-2, -1, 0, 1, 2) to each gene and for each sample. The values give CNA alteration estimation of ploidy status as follows: -2 to homozygous deletion; -1 to hemizygous deletion; 0 to neutral / no variation; 1 to gain; 2 to high level amplification. The genes and the samples were then organized according to the chromosomal aberration of chromosomes 1 and 16 (arm p and q as well) and refers to the 1,16 chromogroups (Privitera et al., 2021) (Table 6). In particular, for chromosome 1 we have 1188 genes on arm 1p, 1078 genes on arm 1q, while for chromosome 16 we have 506 genes on arm 16p, 398 genes on arm 16q. A sum of the of Deletion (∑ -2 and -1), Disomy (∑ 0) and Amplification (∑ +1, +2) for each sample and for chromosomal arms chr1p, chr1q, chr16p, chr16q. Finally, the percentage of chromosomal abnormalities per number of genes of that chromosomal arm was calculated considering the following groups: 1q G, 16q L, 1p L, 16p D/G, 1p D, 16p D, 16q D, 1q D where D = Disomic, L = Loss or deletion, G = Gain or amplification Table 11 shows a capture of the performed calculation to obtain the groups.

	1p loss *n (1p) = 1188*		*1p D* *n (1p) = 1188*		*1q gain* *n (1q) = 1078*		*1q D* *n (1q) = 1078*		*16p D/G* *n (16p) = 506*		*16p gain* *n (16p) = 506*		*16p D* *n (16p) = 506*		*16q loss* *n (16q) = 398*		*16 q D* *n (16q) = 398*	
sample_id	Σ -1 -2	%	Σ 0		Σ 1 e 2	%	Σ 0		Σ 0 1 2	%	Σ 1 e 2	%	Σ 0	%	Σ -1 -2	%	Σ 0	%
MB-0000	0	0	1188	100	2	0.18	1076	99.81	506	100	0	0	506	100	0	0	398	100
MB-0039	0	0	1187	99.91	0	0	1078	100	506	100	0	0	506	100	0	0	398	100
MB-0045	839	70.62	238	20.03	658	61.03	316	29.31	416	82.21	0	0	416	82.21	194	48.74	190	47.73
MB-0046	0	0	1187	99.91	1070	99.25	7	0.64	498	98.41	243	48.02	255	50.39	398	100	0	0
MB-0048	0	0	1187	99.91	173	16.04	904	83.85	449	88.73	138	27.27	311	61.46	96	24.12	302	75.87

Table 11: A sum of the of Deletion (∑ -2 and -1), Disomy (∑ 0) and Amplification (∑ +1, +2) for each sample and for chr1p, chr1q, chr16p, chr16q to obtain 1q G, 16q L, 1p L, 16p D/G, 1p D, 16p D, 16q D, 1q D groups

This is a fundamental step to be able to compare the aforementioned METABRIC cohort with that of the 1,16 chromogroups (Privitera et al., 2021) obtained from TCGA cohort for clinical, prognostic and predictive purposes.

The percentage values were then used to identify chromogroups 1,16 considering a percentage above 80% obtaining 447 samples for Group A, 154 for Group B, 159 for Group C, 183 for Group D and 292 for Control (see Table 12). The samples were further grouped according to intrinsic molecular subtypes according to PAM50: LumA (n = 700); LumB (n = 475); Her2 (n = 224); Normal-like (n = 148); Basal-like (n = 209); Claudin-low (n = 218).

	group A	*group B*	*group C*	*group D*	*CONTROL*
1q G	>80%	>80%	>80%		
16q L	>80%			>80%	
1p L		>80%			
16p D/G	>80%			>80%	
1p D	>80%			>80	>80%
16p D			>80%		>80%
16q D			>80%		>80%
1q D				>80	>80%
TOT	**447**	**154**	**159**	**183**	**292**
LumA (n= 700)	278	36	**34**	**99**	**53**
LumB (n = 475)	88	67	**43**	**42**	**27**
Her2 (n = 224)	16	15	**27**	**6**	**31**
Basal-like (n = 209)	5	6	**23**	**1**	**11**
Normal-like (n = 148)	27	7	**11**	**13**	**42**
Claudin-low (n = 218)	6	7	**8**	**6**	**101**

Table 12: Percentage of chromosomal abnormalities by number of genes in the chromosomal arms considered and classification in intrinsic subtypes.

4.1.11 Overall Survival: Comparison of TCGA vs METABRIC cohorts

The prognostic parameters of Overall Survival (OS), Progression-Free Survival (PFS) and Disease-Free Survival (DFS) in according to 1,16 chromogroups and intrinsic molecular subtype were evaluated. In particular, the available clinical data for TCGA cohort concern OS, PFS and DFS, while for METABRIC cohort, available clinical data refers to OS only. For this reason, in this thesis only the OS analysis will be showed.

The Overall Survival is the gold standard parameter to measuring how long patients with a specific condition (i.e. disease) or after a treatment can survive.

4.1.11 A - TCGA COHORT: intrinsic molecular subtypes

Each of the intrinsic subtypes (LumA, LumB, Her2, Normal-Like and Basal-Like) was compared with the more abundant group of LumA. In **Figure 32**, LumB group shows a worse prognosis than LumA with a significant p-value (p = 0.0140). However, the median survival of the LumB group have no difference respect to LumA group. This is due to a limited number of patients and the follow-up of LumB group. The comparison between LumA vs Her2 also shows invalid results due to the limited number of patients, although the statistical test indicates a significant difference between the two curves (p = 0.0087).

The comparison of the LumA vs Normal-Like and the LumA vs Basal-Like groups do not show statistically significant results (p = 0.1912, and p=0.8105 respectively) due in part to the small number of the other groups in comparison with LumA.

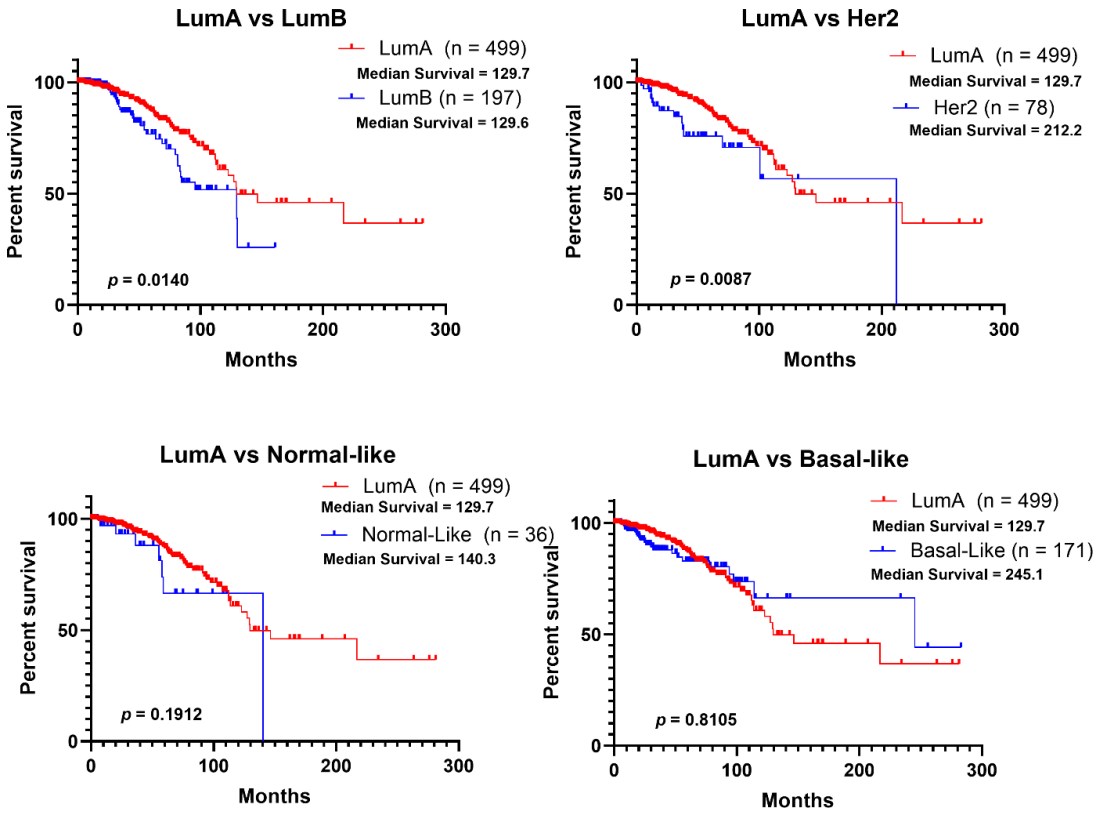


Figure 32: Survival curves for TCGA cohort. Comparisons of intrinsic molecular subtype according to PAM50.

4.1.11 B - TCGA COHORT: 1,16 Chromogroups

The comparison among chromogroup B, C and CONTROL and chromogroup A shows, differences both in median survival values and in statistical significance (**Figure 33**). The chromogroup A show a median survival value twice respect to others chromogroups, thus demonstrating that the prognosis associated with chromogroup A is the most favourable. The Comparison between chromogroup A and C is "undefined" due to the number of live patients exceeding 50 % of the cases considered at the longest follow-up time.

In addition, the comparison between Group A vs B shows how group B is the one that has a worse prognosis even in comparison with the other chromogroups (see **Figure 33**).

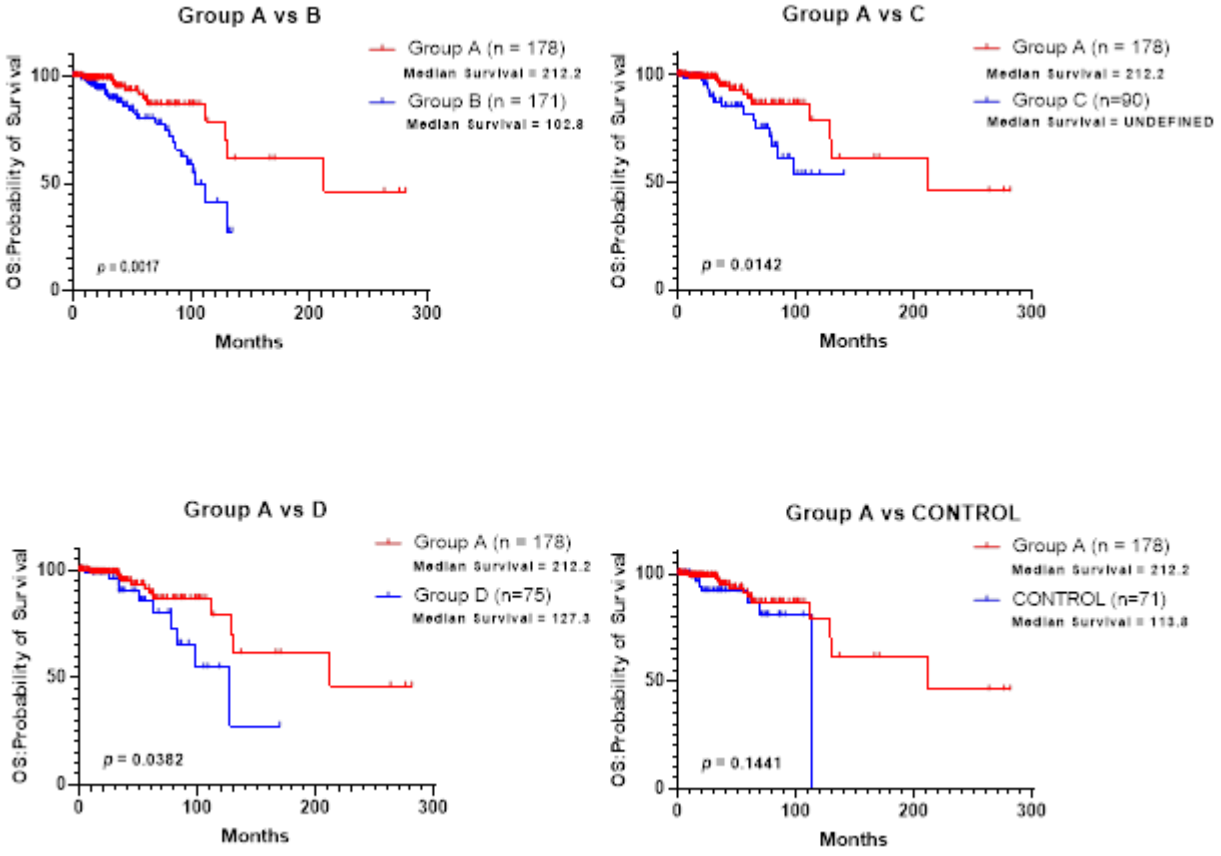


Figure 33: Survival curves with Overall Survival (OS) data from the TCGA Samples divided by chromogroups 1,16.

As for the statistical significances, the p-value values are significant for all the chromogroups compared with group A, except when comparing group A with the CONTROL group.

In conclusion, the data do not provide a definitive result due to the small samples and duration of follow-up and the need to be confirmed in a bigger cohort such as the METABRIC one.

4.1.11 C - METABRIC COHORT: Intrinsic molecular subtypes

The METABRIC cohort was subdivided according to intrinsic molecular subtypes LumA, LumB, Her2, Normal-like, Basal-Like and Claudin-low. The prognosis of LumB, Her2 and Basal-Like groups is worse compared to LumA which is confirmed to be the one with favourable prognosis after diagnosis (**Figure34**).

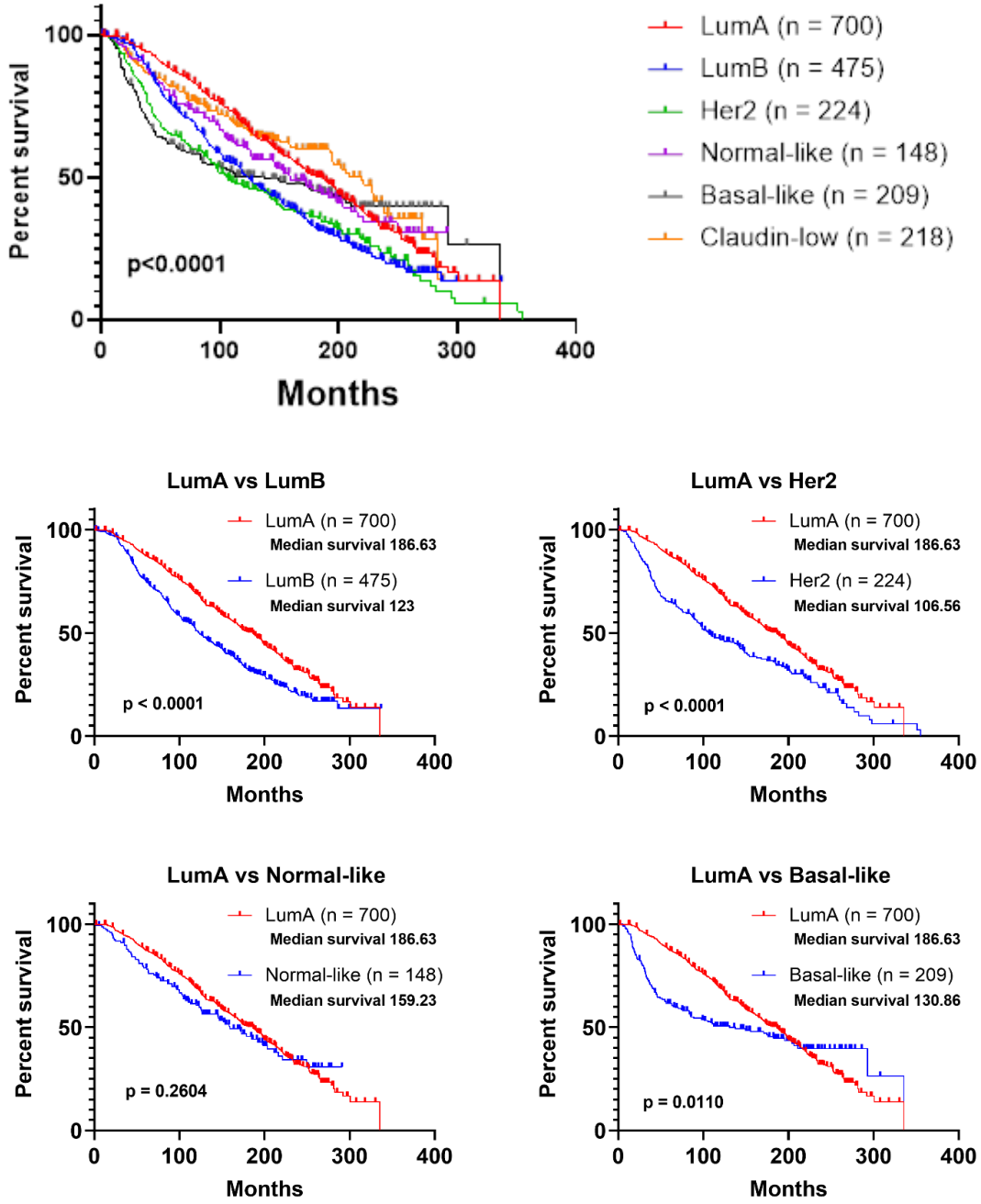


Figure 34: Survival curves of the Metabric cohort of samples by intrinsic molecular subtype. OS all subtypes (up panel). Comparisons of LumA vs all other intrinsic subtypes.

These differences are confirmed both by more suitable Median Survival values (see Table 13) for LumA (Median Survival = 186.6 months) compared to the LumB (Median Survival = 123

months), Her2 (Median Survival = 106.6 months) and Basal-like (Median Survival = 130.9 months) groups and by the significant differences given by the log-rank test for LumA vs LumB ($p < 0.0001$), LumA vs Her2 ($p < 0.0001$) e LumA vs Normal-Like ($p = 0.011$) (Table 13).

INTRINSIC MOLECULAR SUBTYPES	MEDIAN SURVIVAL
LumA	186.6
LumB	123
Her2	106.6
Normal-Like	159.2
Basal-Like	130.9
Claudin-Low	219.2

Log-Rank Test p-value	
LumA	
<0.0001	**LumB**
<0.0001	**Her2**
0.2604	**Normal-Like**
0.0110	**Basal-Like**
0.4089	**Claudin-Low**

Table 13: **Median survival values for each intrinsic molecular subtype (left panel); Log-Rank Test p-value LumA (right panel) compared with all other intrinsic molecular subtypes (METABRIC cohort).**

4.1.11 D - METABRIC COHORT: 1,16 Chromogroups

Analysis of the survival curves of chromogroups 1,16 (**Figure 35**) shows that Group A has a better prognosis than chromogroups B and C and overlaps with both group D and the Control group. These differences are confirmed both by the better median survival values for groups A (Median Survival = 180.8 months), D (Median Survival = 184.7 months) and Control (Median Survival = 199 months) compared to groups B (Median Survival = 112.6 months) and C (Median Survival = 128.5 months), and by the Log-Rank Test (see table 14).

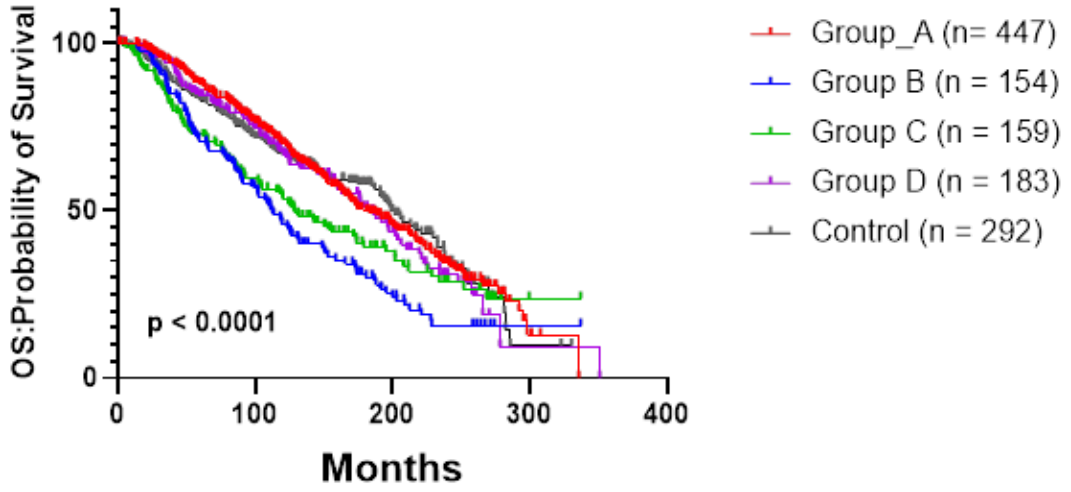


Figure 35: Survival curves of the Metabric cohort of 1.16 chromogroups

Cromogruppi 1,16	MEDIAN SURVIVAL
Group A	180.8
Group B	112.6
Group C	128.5
Group D	184.7
Control	199

Log-Rank Test p-value	
Group A	
<0.0001	**Group B**
0.0080	**Group C**
0.5307	**Group D**
0.9858	**Control**

Table 14: **Median survival values for each 1,16 chromogroups (left panel); Log-Rank Test p-value LumA (right panel) compared with all 1,16 chromogroups (METABRIC cohort).**

However, the differences are significant only for comparisons between Group A vs B ($p < 0.0001$), and Group A vs C ($p = 0.0080$). Even in the METABRIC cohort group B has a worse prognosis than all other chromogroups considered with a low survival median of 112.6 months.

4.1.11 E - Diagnostic value of Group A

A further survival analysis was performed considering the portion of group A samples with better prognosis and group B with worse prognosis for the intrinsic molecular subtypes LumA and LumB. The purpose of this survival analysis is to compare whether the new molecular cytogenetic classification in chromogroups 1,16 may represent an alternative method to assess prognosis compared to the already used classification system in intrinsic molecular subtypes (**Figure 36**).

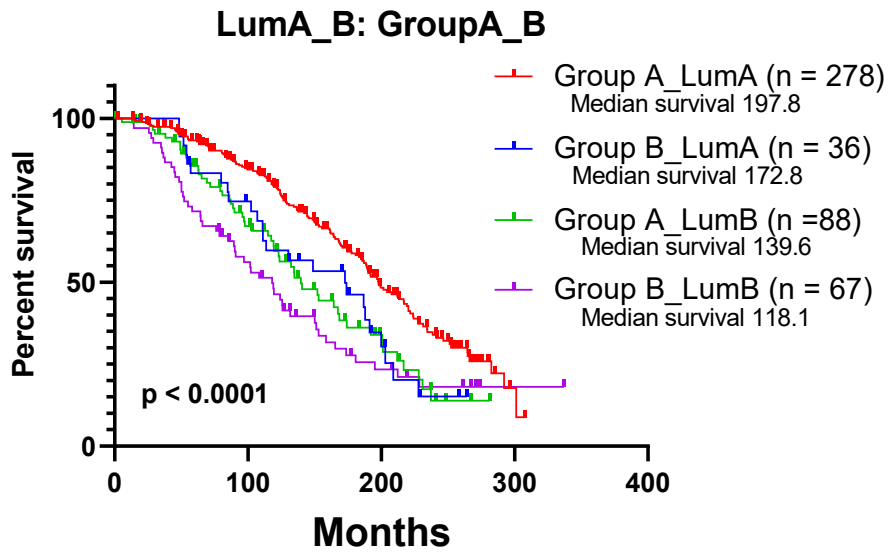


Figure 36: Survival curves of the Metabric cohort of samples divided by intrinsic subtypes LumA in chromogroups A and B and LumB in chromogroups A and B.

The Group A – LumA is the group with a more suitable prognosis than the following groups: Group B-LumA, Group A-LumB and Group B-LumB with a median survival of 197.8 months and, as shown in Table 15, with a statistically significant difference.

Log-Rank Test p-value				
	Group A_LumA	**Group B_LumA**	**Group A_LumB**	**Group B_LumB**
Group A_LumA	//	**p = 0.018**	**p = 0.0003**	**p < 0.0001**
Group B_LumA	p = 0.018	//	p = 0.8170	p = 0.2277
Group A_LumB	p = 0.0003	p = 0.8170	//	p = 0.2340
Group B_LumB	<0.0001	p = 0.2277	p = 0.2340	//

Table 15: p-value given by the Log-Rank Test of the intrinsic subtypes LumA in chromogroups A and B and LumB in chromogroups A and B (METABRIC cohort).

We can therefore deduce that the molecular cytogenetic classification in chromogroups 1,16 analysed in both the TCGA and METABRIC cohorts, could represent a useful and alternative method to evaluate the prognosis in cases of invasive breast cancer as an alternative or in addition to the classification in intrinsic molecular subtypes.

4.1.12 ETV6-NTRK3 gene-fusion

Secretory breast cancers (SBCs), as described in par. 1.6.2, are a very rare phenotype characterized by distinctive genomic pattern with a recurrent chromosomal translocation t(12;15)(p13;q25), leading to *ETV6–NTRK3* fusion gene.

In order to explore the expression profiling associated to ETV6-NTRK3 fusion gene, we analysed TCGA cohort categorizing the carcinoma samples by using the transcript per million (TPM) values and the corresponding prognostic data. In particular, a threshold was determined to individuate an expression value, above which NTRK3 gene is up-regulated. The average of TPM (1.22 AvgTPM) and the standard deviation (3.08 SD) for the NTRK3 gene were calculated. The critical threshold was set up at 7.38 (AvgTPM + 2-fold standard deviation) obtaining 32 samples NTRK3 TPM> 7.38 and 1026 NTRK3 TPM < 7.38. The subclassification of these main groups in intrinsic subtype and 1,16 chromogroups is showed in table 16.

The expression data so organized were matched with the corresponding prognostic data. The overall survival (OS) was valuated through Kaplan-Meier curve (**Figure 37**).

	group A	group B	group C	group D	CTRL
	TCGA (n = 1058) - RNAseq				
Total	**175**	**165**	**89**	**72**	**68**
NTRK3 TPM > 7.38 (n = 32)	**4**	**1**	**2**	**3**	**4**
LumA (n=11)	3	0	0	3	0
LumB (n = 4)	0	1	0	0	1
Her2 (n = 1)	0	0	1	0	0
Basal-like (n = 10)	0	0	1	0	0
Normal-like (n = 1)	0	0	0	0	1
NTRK3 TPM < 7.38 (n=1026)	**171**	**164**	**87**	**69**	**64**
LumA (n=479)	132	78	29	57	33
LumB (n =188)	17	53	27	8	5
Her2 (n =76)	4	15	8	1	5
Basal-like (n = 158)	0	7	18	0	2
Normal-like (n =34)	2	0	3	2	12

Table 16: TCGA Cohort (1058 RNA-seq samples) subdivided in 1,16 chromogroups (group A-B-C-D and control) and intrinsic molecular subtype.

In a study involving six women aged between 25 and 60 year with secretory cancer, the rearrangement ETV6-NTRK3 fusion gene was evaluated. In particular samples were tested both with interphase fluorescence in situ hybridization (FISH) methodology to verify the presence of the rearrangement and with immunohistochemistry to evaluate the histological classification (Laé et al., 2009). Their results classified secretory breast carcinoma as triple-negative expressing basal markers (Laé et al., 2009). In comparison to other cancer types such as basal-like breast carcinomas and adenoid cystic carcinoma, secretory breast carcinomas show a favourable prognosis in their series. (Laé et al., 2009).

For these purposes, NTRK3 > 7.38 TPM (NTRK+) group was compared with NTRK3 < 7.38 TPM (NTRK-) group (**Figure 37A**). Subsequently they were compared with Basal NTRK+/- (**Figure 37B**) and Luminal A NTRK3 +/- groups (**Figure 37C**).

In **figure 37A**, NTRK3+ group shows a more favourable survival rate consisting of more than 50% alive samples at the longest follow-up time (median survival undefined) respect to NTRK3- group with a median survival of 129,1 (Log-Rank test significance 0.024). In **figure 37B** the comparisons were performed among NTRK3+, NTRK3+ Basal and the control group NTRK3-. The results not only confirm that NTRK3+ have a better prognosis compared to the entire TCGA cohort, but they show a favourable prognosis also in the context of basal subtype (Log-Rank test significance 0.0169). As shown in table 15, the most abundant groups both in NTRK3+ than NTRK- are Luminal A and Basal. In **figure 37C** the NTRK3+/- LumA and Basal and NTRK3+/- LumA and Basal were compared.

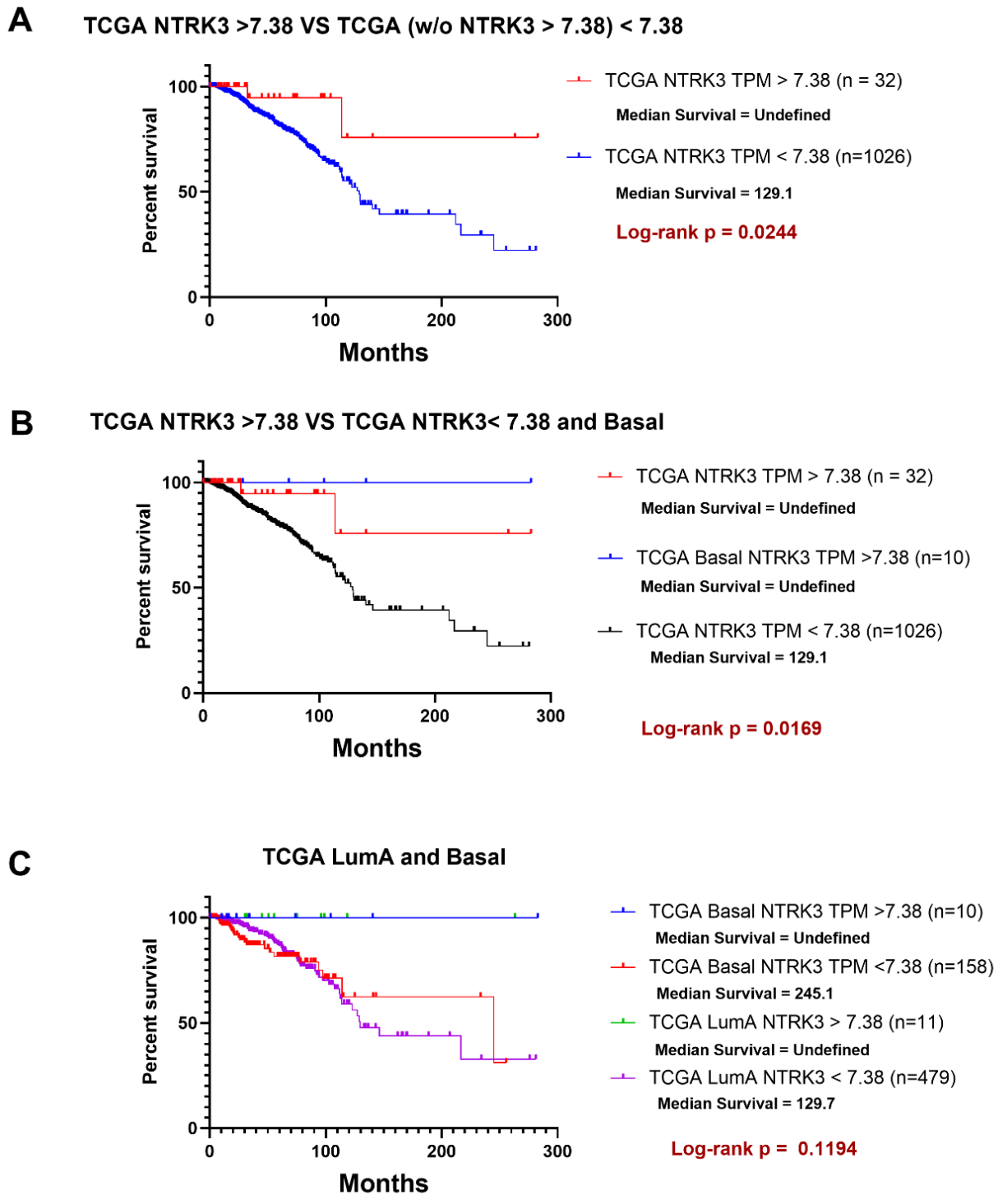


Figure 37: Kaplan-Meier of NTRK3+ and NTRK3.

The groups NTRK3+ both Basal and LuminalA show an overlapping survival curve (Blu and green curves) in comparison to NTRK3- (Basal and LuminalA) groups. These results confirmed not only the favourable prognosis but also, they suggest the possibility to have a specific gene marker to perform better diagnosis of the secretory carcinoma.

4.1.13 Choosing cell model: CAL148 – CAL-51

In order to translate the bioinformatic model of 1,16 chromogroups in a suitable cell models, an additional bioinformatics analysis was performed. From "Cell Model Passport" database we downloaded data on 978 cell lines, by which 80 were breast cell lines. Only 52 cell lines, out of 80, had transcriptomic (RNAseq) and cytogenetic (SNParray) data. The cytogenetic data, previously analysed with Gistic2 algorithm (Mermel et al., 2011) (Beroukhim et al., 2010) (Beroukhim et al., 2007) have assigned a numeric value (-2, -1, 0, 1, 2) to each gene entries in the genome references and for each sample. The number of genes for each chromosomal arm of interest was then calculated. In particular, for chromosome 1 we have 1121 genes on arm 1p, 1006 genes on arm 1q, while for chromosome 16 we have 483 genes on arm 16p, 378 genes on arm 16q. The sum of Deleted genes (∑ -2 and -1), diploid genes (∑ 0) and Amplified genes (∑ +1, +2) for chromosomal arm (chr1p, chr1q, chr16p, chr16q) was calculated for each cell line. The only two cell lines responding to the der(1,16) and the model not bearing aberrations of Chr1 and Chr16 were CAL-148 and CAL51, respectively. As shown in table 17, CAL148 bear 1q-gains and 16q-loss while CAL-51 cells are disomic for chromosome 1 and chromosome 16.

		SIDM00933	***SIDM00938***
	∑(-2, -1, 0, 1, 2)	**CAL-51**	**CAL-148**
1p	Value -2	0	0
	Value -1	0	**648**
	Value 0	**1121**	**446**
	Value 1	0	26
	Value 2	0	1
n = 1121			
1q	Value -2	0	1
	Value -1	0	1
	Value 0	**995**	3
	Value 1	0	**922**
	Value 2	11	**79**
n=1006			
16p	Value -2	0	0
	Value -1	0	1
	Value 0	**483**	3
	Value 1	0	**479**
	Value 2	0	0
n=483			
16q	Value -2	0	**1**
	Value -1	0	**377**
	Value 0	**378**	0
	Value 1	0	0
	Value 2	0	0
n=378			

Table 17: Gistic analysis and sum of corresponding values (1p, 1q, 16p,16q) in CAL148 and CAL51.

To validate the CAL148 and CAL51 cell models as well as to verify if the gene expression profiles of these cell lines agreed to 1,16 chromogroups gene pattern previously described, we performed a differential expression analysis. The first step consists in the application of a mathematical method to obtain the 1,16 chromogroup from GISTIC results. We calculated the percentage of altered genes per chromosomal arm, establishing that the arm was in gain when more than 90% of genes, located in that arm, had a value 1 or 2, and the arm was in loss when more than 90% of gene had a value -1 or -2. For disomy the arm had more than 90% of gene with 0 value.

1qG/16qL	CTRL (1qD/16qL)
1q Gain > 90 %	1q Disomic > 90 %
16q Loss > 90%	16q loss > 70%
BT-549 HCC2218 HCC2157 HCC1599 HCC1569 CAL-148 EFM-19	Hs-578-T MDA-MB-231 MDA-MB-330 HCC1419

Table 18. Selected cell groups 1qG/16qL CTRL (1qD/16qL).

Due to the small number of cell lines responding to our parameters we considered only two groups 1qG/16qL and group Control (CTRL). The CTRL group was characterized by disomic 1q and 16q loss, in this case a loss in >70% of genes was considered enough to reveal arm chromosome loss. The group 1qG/16qL is similar to group A and B1, while the group 1qD/16qL is similar to group D in 1,16 chromogroups classification. This filtering allowed us to include into the 1qG/16qL group 7 cell lines (BT-549, HCC2218, HCC2157, HCC1599, HCC1569, CAL-148, EFM-19) while to include into the CTRL (1qD/16qL) group 4 cell lines (Hs-578-T, MDA-MB-231, MDA-MB-330, HCC1419) (Table 18).

4.1.13 A - Transcriptomic Data

The transcriptomic data relative to cell lines are also retrieved from the Cell Model Passport repository. In particular, the RNAseq data covers about 1047 cell lines and about 36203 genes. Only 52 cell lines, which had the transcriptomics (RNAseq) and cytogenetics (SNParray) data were available. Transcripts are classified according to their differential expression between the 1qG/16qL group (group under study) and the 1qD/16qL group (which is considered as a control group). If the expression of a transcript is higher in 1qG/16qL than in control group (linear fold change vs Control >1.5) the transcript is called **OverT** (Overexpressed Transcript). The chromosomal distribution of OverT, normalized for the

number of genes in each chromosome arm, has been determined by calculating the so called normalized chromosomal distribution index (NCDI).

For comparison the NCDI has been calculated also for all transcripts, independently by their expression status: since the value is normalized for gene chromosomal content all the chromosomal arms show an NCDI value equal to 2.38% (100 / 42 chromosomal arms) (blue columns in **Figure 38**). On the contrary the NCDI values for OverT show remarkable differences among chromosomal arms (green columns in **Figure 38**). NCDI value for OverT is 4.5% in chromosome arm 1q. This means that 4.5% of all overexpressed transcripts (OverT) are localized in the chromosomal arm 1q. Since gene copy number in 1q is increased, this result confirms the presence of a gene-copy effect (upregulation of transcription due to an increased gene copy number). The NCDI values of OverT transcripts in 16p and the 16q do not show any differences respected to all transcripts, in agreement with the fact that the study group (1qG/16qL) and the control group (1qD/16qL) have the same gene copy number at the level of chromosome 16.

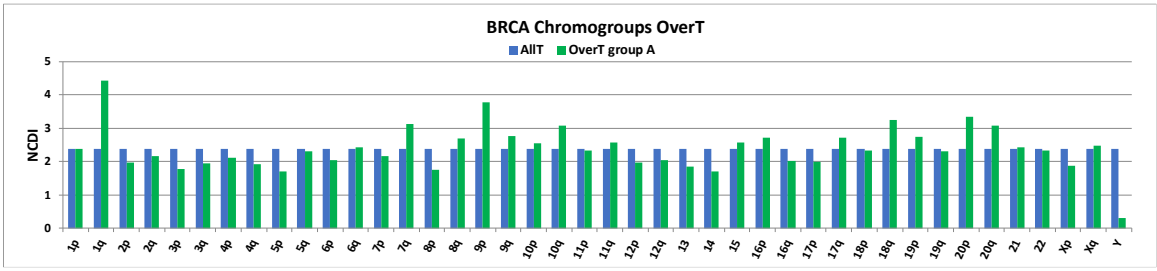


Figure 38 The normalized chromosomal distribution index (NCDI) of OverT of 1qG/16qL. NCDI values of all transcripts (AllT) analysed by RNA-seq in each chromosomal arm are also reported for comparison.

Subsequently, we verified if the genes over-expressed in 1q in cancer cell cultures of the 1qG/16qL group are the same genes overexpressed in tumour cells in vivo. This is not an obvious result since the artificial culture conditions deeply affect gene transcription and only genes strictly related to the cell phenotype keep their expression level. Therefore, we compared the list of OverT genes localized on chromosome 1q identified in the present study with a list of 639 OverUpT genes (overexpressed in comparison to control cancer and upregulated in comparison to normal tissue) obtained in a comparison between breast cancer belonging to chromogroup A (bearing 1q gain) with the cancer control group (not bearing 1q gain) (Table S2 from Privitera et al 2021). Interestingly, 251 (out of 639 genes) identified in tumoral samples were also overexpressed in breast cancer cell cultures (Table 19).

Gene Set Enrichment Analysis was previously used to find genes located in 1q with significant functional correlations in carcinogenesis and cancer progression (Privitera et al 2021). On the basis of those results we analysed the specific expression of genes related to the

beta-catenin pathway (BCL9, PYGO2), genes encoding for the subunits of γ-secretase complex (APH1A, PSEN2 and NCSTN), and genes CDH1A, BCL9 as well as EFNA3 and the EFNA1, which encode a class of protein that interact with receptor Eph. These genes were selected in accordance with the results in Privitera et al. 2021, which have demonstrated how these genes are important in breast cancer tissue bearing derivative 1,16 (Group A).

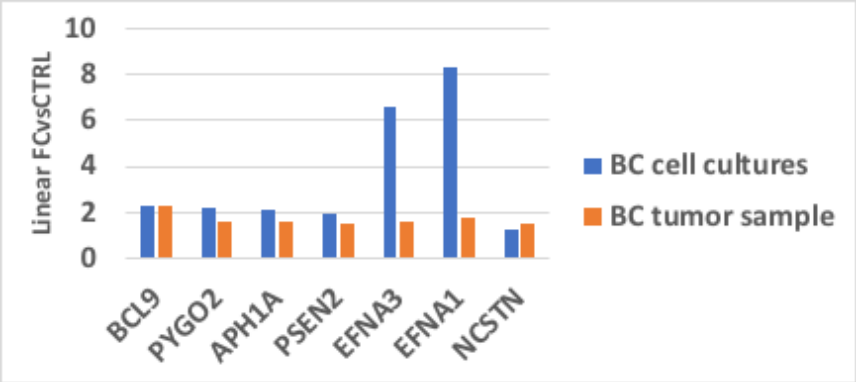


Figure 39. Comparison of gene-expression levels between Linear fold change vs Control (linear FCvsCTRL) of cell cultures (Group 1qG/16qL) and the tumour sample of Group A from (Table S2 from Privitera et al 2021).

List of 251 over-expressed genes located in chromosome 1q				
DESI2	OR3D1P	GBAP1	ARV1	CREB3L4
KDM5B	PIP5K1A	ZNF687	OR14K1	PARP1
PPOX	COA6	GAS5-AS1	SSR2	MRPL55
DAP3P1	DPM3	PFN1P4	C1orf21	RNU6-884P
C1orf35	LYSMD1	GOLT1A	RRP15	RNY4P25
VPS45	KIF28P	PFKFB2	GATAD2B	KCTD3
POGK	GNPAT	TFB2M	TMEM183A	ENAH
ISG20L2	RAB25	TARS2	MSTO1	PSMD4
FBXO28	STX6	RD3	NTPCR	BOLA1
GPR89B	PLEKHA6	RBM34	OAZ3	ENSA
BRINP3	MUC1	ASH1L	RRM2P2	RNA5SP78
IL20	ANGEL2	MSTO2P	DUSP12	SLC50A1
F11R	NVL	GGPS1	URB2	FDPSP8
TRIM46	SOX13	YOD1	NR1I3	GAS5
PRCC	THEM4	TOR3A	UCHL5	COG2
TSACC	PIGM	NDUFS2	B3GALNT2	IARS2
GON4L	CNIH4	UBQLN4	MAPKAPK2	FAM189B
SCAMP3	IRF6	SPRTN	DNAH14	TADA1
SNRPE	HAX1	CLK2	TOMM20	C1orf147
RUSC1	ASPM	CHTOP	RNF115	GABPB2
LAMTOR2	INTS3	KCNH1	VANGL2	OR1C1
DAP3	SELENBP1	PRPF3	SLC39A1	HNRNPA1P59
CCSAP	UBE2Q1	APH1A	C1orf43	TUFT1
SETDB1	POLR3C	ZNF496	IPO9	TMEM79
RAB4A	C2CD4D	RN7SL444P	MTX1	TDRKH
TDRD5	MRPL24	ZNF692	F13B	ADAR
OR6F1	ZNF695	ADAM15	POU5F1P4	ATP1B1
USP21	FH	FMO9P	LEFTY1	MROH9
USF1	PKLR	RPRD2	LIN9	EFNA1
INTS7	TP53BP2	OCLM	DCAF8	C1orf189
SNAP47-AS1	SNX27	DARS2	ZNF669	RIIAD1
FLAD1	HSPA6	GOLPH3L	NBPF11	TARBP1
BPNT1	MDM4	RUSC1-AS1	MEX3A	MIR4258
SERTAD4	POGZ	DCST1	RPS27	DUSP23
KLHDC9	TMEM9	UBAP2L	PSEN2	CCT8P1
NUP133	TPR	LRRN2	GPATCH4	ESRRG
CENPF	TOMM40L	TTC13	B4GALT3	PAQR6
FAM72A	PEX11B	ARHGEF11	ABCB10	JTB
BCL9	TAF5L	FCGR3B	TSNAX	DEGS1
CRABP2	OTUD7B	KRTCAP2	PYGO2	IKBKE
CRTC2	LENEP	HDGF	SLC25A44	PMVK
HEATR1	C1orf131	RFX5	YY1AP1	FLVCR1
PGBD2	ZBTB7B	AGT	LINC01136	PI4KB
SCNM1	SRP9	HSD17B7	ANP32E	ASH1L-AS1
ZNF672	SNAP47	RCOR3	EFNA4	DCST2
SCCPDH	IGSF9	VPS72	ZNF124	BTG2
TRIM11	C1orf53	UFC1	KCNH1-IT1	MRPL9
METTL18	LYPLAL1	SDHC	CDC42SE1	USH2A
TMEM63A	LHX4	MRPS14	MTX1P1	RBBP5
AHCTF1	EFNA3	IL19	GBA	CKS1B
				SF3B4

Table 19. List of 251 genes located in chromosome 1q and overexpressed in breast cancer cell cultures bearing 1qG/16qL and in tumoral samples belonging to chromogroup A when compared to control cancer not bearing 1qG.

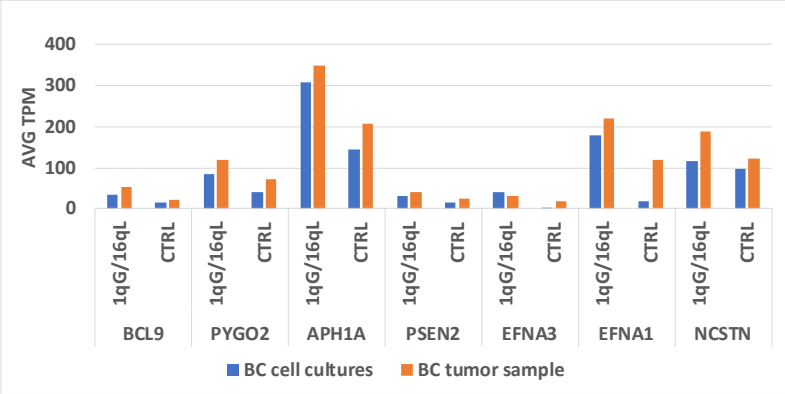


Figure 40. Comparison of TPM levels between cell cultures (Group 1qG/16qL) and the tumour sample of Group A from (Table S2 from Privitera et al 2021).

The expression levels of such genes in cell cultures and in tumoral samples was compared. In **figure 39** the expression level is reported as linear fold change vs Control (linear FCvsCTRL). As shown in **figure 39**, all selected genes show an increased expression ranging from 1,5 to 8-fold increase. The higher expression is observed for EFNA3 and EFNA1 genes in cancer cell cultures. **Figure 40** shows the same results expressed as TPM (transcripts per million). In **Figure 40** it is possible to observe the values in control tumours or in control cultures. In agreement with the fold-change values reported in **Figure 39**, TPM values are higher in 1q-bearing cells in comparison to control cells. However, it is interesting to note that the dramatic differences in fold-change in EFNA3 and EFNA1 genes in cell cultures are mainly due to a reduced expression of these genes in control culture. Since tumoral samples include stromal cells, it is possible that the difference between results in culture and in vivo may depend by the presence of stromal cells. These results confirm the experimental design of the cells belonging to the group A.

4.1.14 Growth curve: CAL-148; CAL51; MCF7

In order to assess and to evaluate the cell culture growth and their maintenance, different experiments were performed with CAL148, CAL 51, MCF7 cell lines.

1) Cal-148, CAL-51 +/- EGF; 2000, 4000, 8000, 16000 cells/well concentration, at four time points (24, 48, 72, 144 h). The aim was to evaluate the different growth rate at different cells/well plating density as well as the effect of the EGF on CAL148.

The **figures 41-42** shows a massive effect of EGF on CAL148 at all cell concentrations and timepoints. The CAL148 EGF+ at 16.000 cells/well shows a negative trend due to a fast growing and the space limitation of well bringing the cell to death. The same effect at 16.000 cells/wells is observable for CAL51 and MCF7. The CAL51 shows a high growth rate as well as a strong and resistant phenotype.

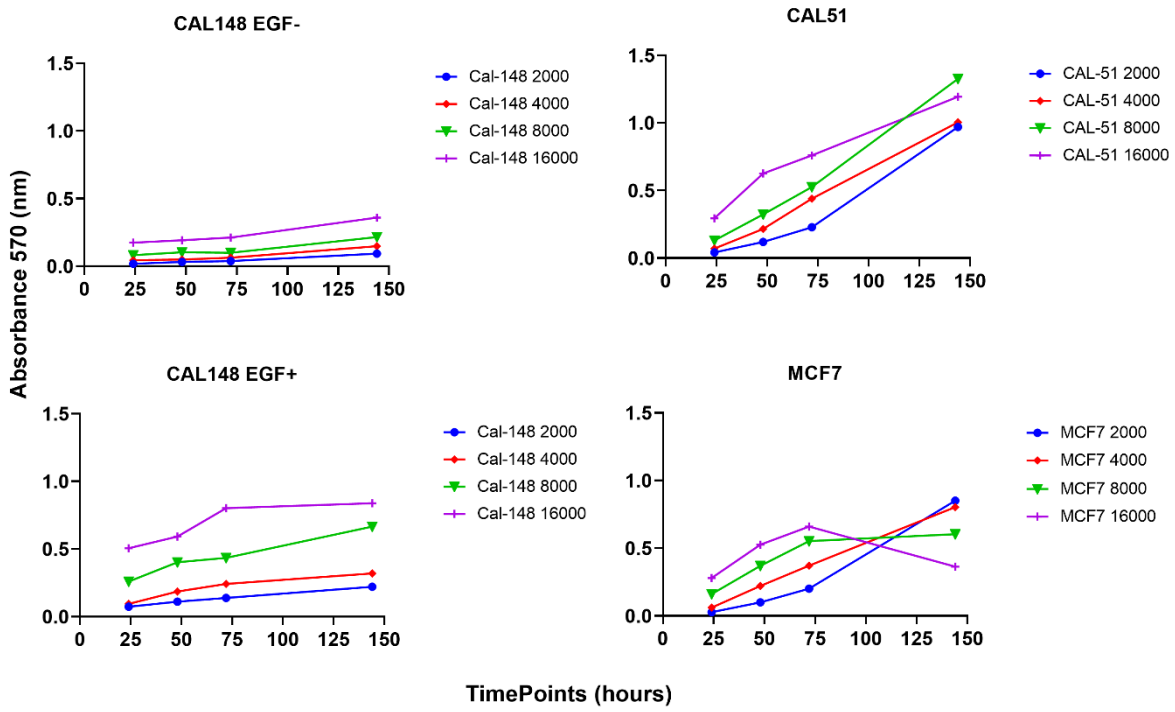


Figure 41: Growing Curve of CAL148 EGF+/-, CAL51, MCF7 at different timepoint and cells/well concentration.

In **figure 42** the effect of EGF on CAL148 was valuated. Unpaired t-test was used to verify the statistical significance for the groups of treated (EGF+) and un-treated (EGF-) for each cell concentration (2000, 4000, 8000, 16000 cells/well) and timepoints (24h, 48h, 72h,144h) in CAL148. In addition, the significance was also tested with non-parametric Mann-Whitney test (Sum rank test) and for all concentrations at any timepoints. The results were overlapping with the results of unpaired t-test.

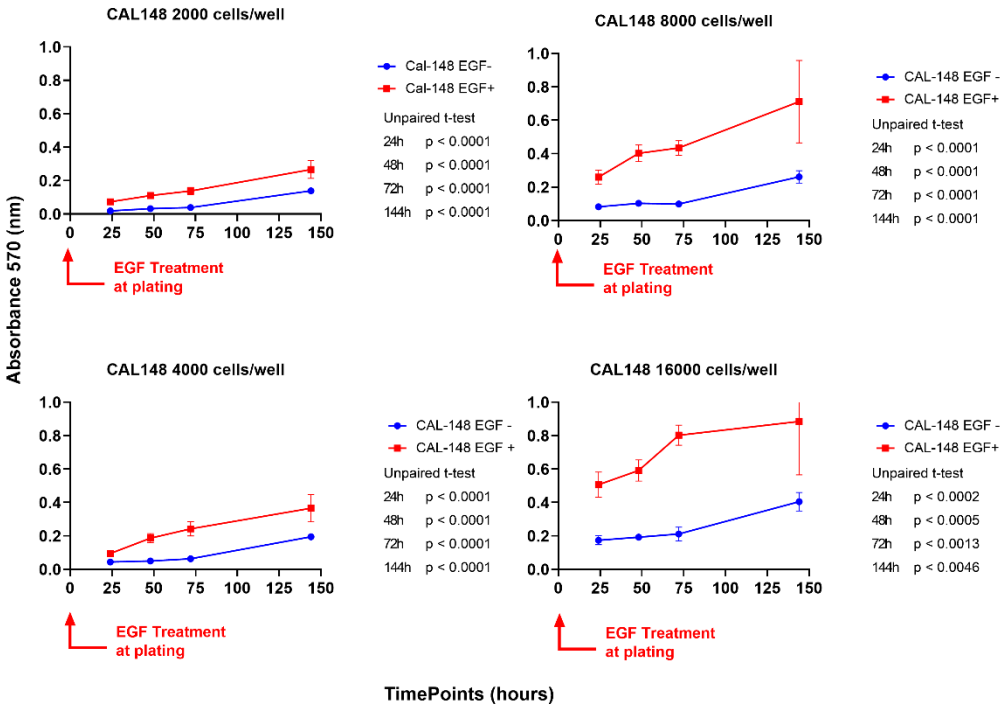


Figure 42: Effect of EGF on CAL148. Unpaired t test was performed at any timepoints. EGF treatment was performed at plating.

There is a statistically significant difference between CAL148 EGF+ and CAL148 EGF- in all conditions. The EGF effect does not disperse over time. EGF was added at the cell plating.

2) Cal-148, CAL-51 and MCF7, +/- EGF treatment, at 4000 and 8000 cells/well density, at five timepoints (24, 48, 72, 96, 144). The EGF was added 24h after plating. The aim was to evaluate the EGF effects only after cell adhesion as well as to identify which cell lines are more sensitive to the treatment.

The results, showed in **figure 43**, demonstrate EGF effects on CAL148 24h after treatment. The CAL148 because of their peculiar capacity to grow up in multilayer can grow much more in a single well-plate in comparison to other cell types. CAL148 are stimulated to grow constantly. The greatest peak of growth is observed between 96h and 144h hours both at 4000 and 8000 cells/well in CAL148. EGF exerts its effect 24h after addition (48h after plating as shown in the graphs) with appreciable and significant differences from 48h to 144h in CAL148 4000 cells/well (Un-paired t-test 48h p value = 0.0070, 72h p value = 0.0092, 96h p

value = 0.0018, 144h p value = 0.0004) and from 72h to 144h in CAL8000 cells/well (Un-paired t-test 72h p value = 0.0067, 96h p value < 0.0001, 144h p value = 0.0003). In CAL51 and MCF7 the differences are not appreciable and significant showing that these cells do not have a real benefit from the addition of EGF which does not act either at the level of adhesion or growth. In CAL51, the blue curve (EGF-) and red curve (EGF+) show a trend inversion with the un-treated curve higher than treated curve.

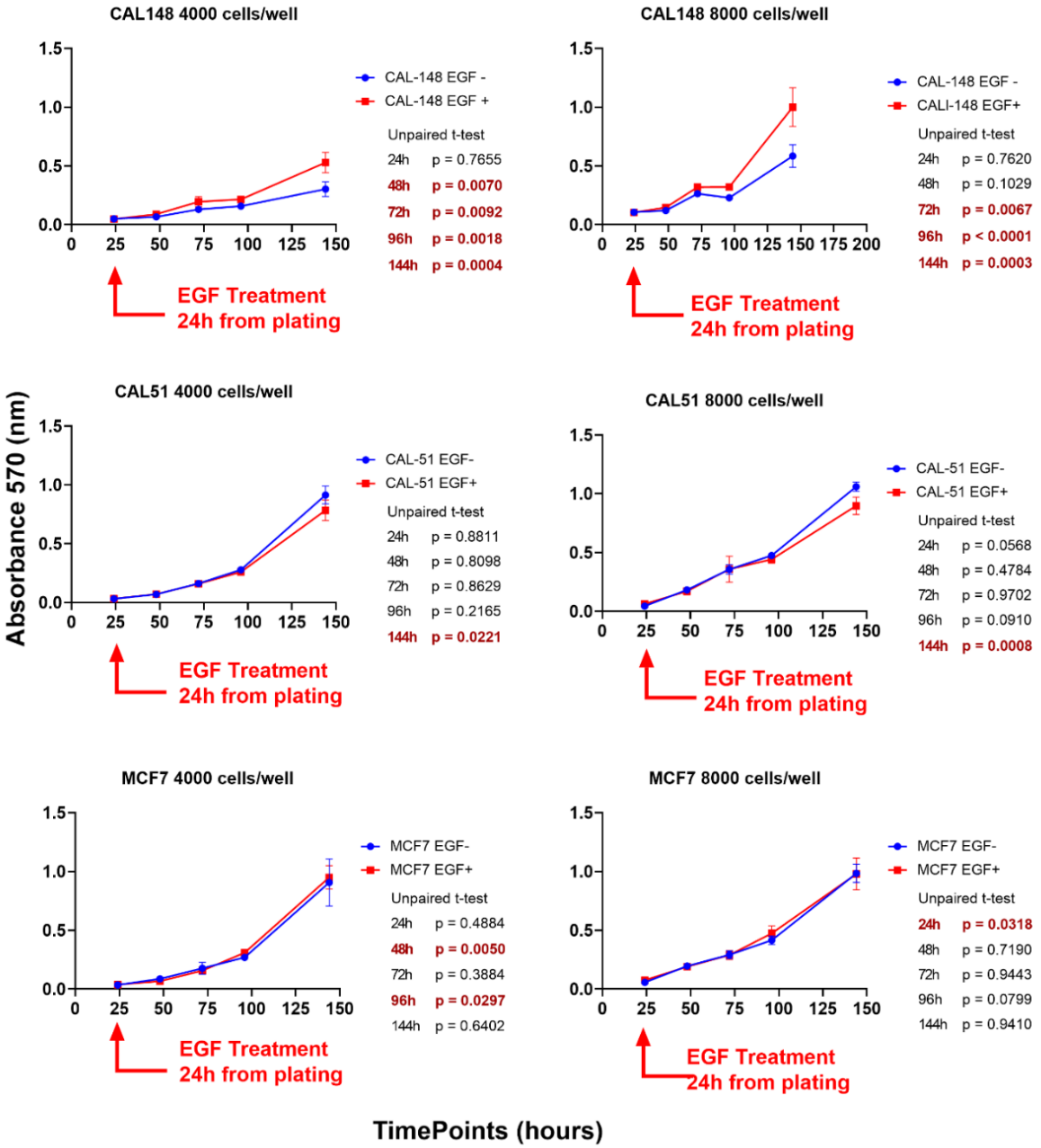


Figure 43: Comparison of EGF+ and EGF- treatments on CAL148, CAL 51, MCF7 with cell concentration of 4000 and 8000 Cells/well at 24, 48, 72, 96, 144 hours from plating. EGF addition at 24h from plating after adhesion. Unpaired t test was performed at any timepoints.

3) Cal-148, CAL-51, +/- EGF treatment 2000, 4000 and 8000 cells/well density, at four timepoints (24h, 48h, 72h, 144h). Two conditions were considered: ***A)*** EGF not pre-treated (1-week) before the assay and treated with EGF (EGF+) or not treated (EGF-) in the assay ***B)*** EGF pre-treated (1-week) before the assay and treated with EGF (EGF+) or not treated (EGF-) in the assay. The data concerning EGF pre-treatment (B) are not showed because they confirm the results in **figures 42 and 43**, while the data concerning experiment A are showed in **figure 44**. The aim was to evaluate the EGF effects before and after the EGF treatment in comparison between CAL148 e CAL51.

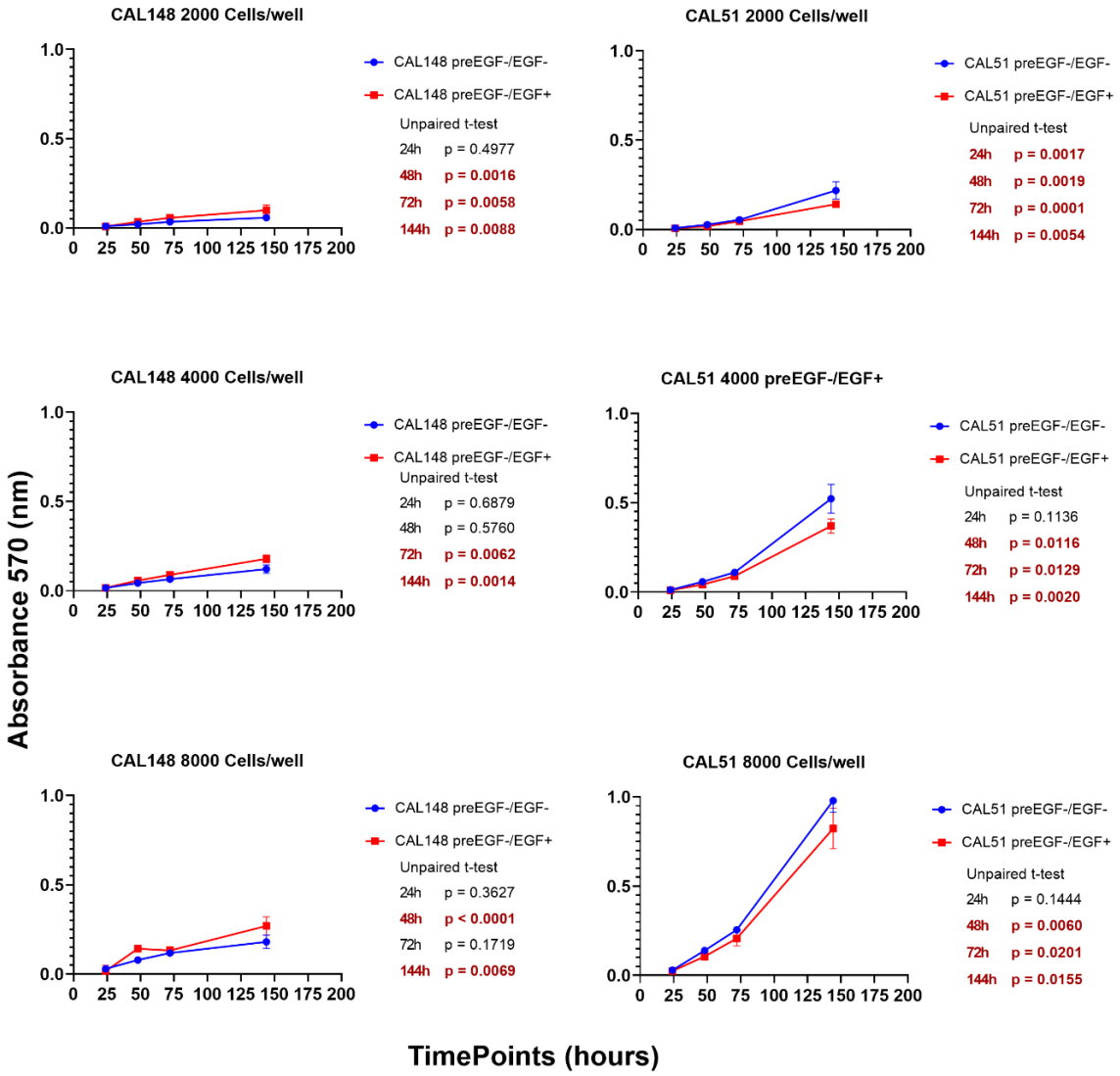


Figure 44: Comparison of EGF+ and EGF- treatments in the not-pre-treated EGF cell population of CAL148, CAL 51 with cell concentration of 2000, 4000 and 8000 Cells/well at 24, 48, 72, 144 hours from plating. EGF addition at 24 plating. Unpaired t test was performed at all timepoints.

In CAL148 the EGF effects is slower respect to the previous experiments. The statistical significance (tested with un-paired t-test) is observed only after 48h from plating in CAL148. In this group of curves (left panel), the trend seems rather flattened, this is only due to the use

of the same scale of magnitude for the Y axis. In CAL51 curves show an inversion, with the curve of the untreated (blue lines) that exceeds in absorbance the curve of EGF treated (red lines), confirming that EGF does not determine an appreciable advantage as in CAL148. The EGF appears to have an opposite effect on CAL51 and the higher growth of EGF untreated cells is statistically significant at any cell plating density and timepoints.

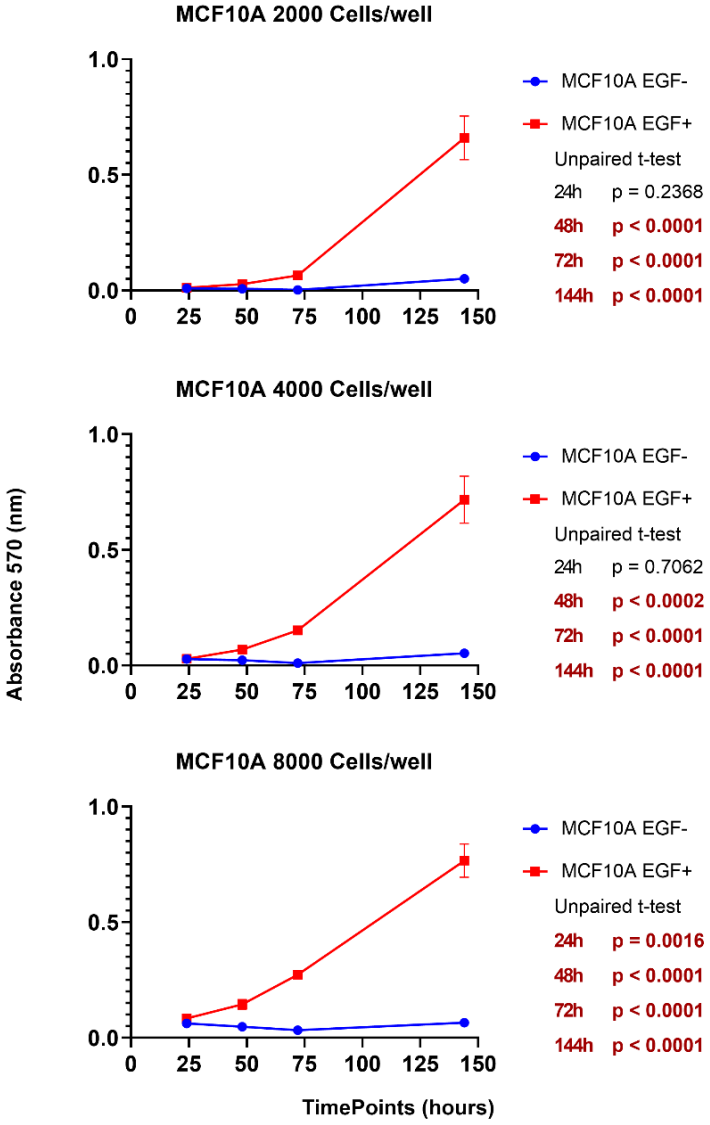


Figure 45: Comparison of EGF+ and EGF- treatments in MCF10A with cell concentration of 2000, 4000 and 8000 Cells/well at 24, 48, 72, 144 hours from plating. EGF addition at plating. Unpaired t test was performed at any timepoints.

4) MCF10A, +/- EGF treatment, 2000, 4000 and 8000 cells/well plating density, at four timepoints (24h, 48h, 72h, 144h). EGF was added at plating. The **figure 45** shows the importance of the EGF treatment on MCF10A. In the cell population treated with EGF (EGF+), the cell growth rate is very high compared to the untreated population with a pinch

of the growth curve already at 48h. The differences are statistically significant at any cell concentrations and timepoints.

These experiments have been a guide to determine the best cell density and how to ensure the supply of EGF in subsequent experiments. We concluded that the best average cell density is at 4000 cells/well for CAL148, CAL51, MCF7 and MCF10A. This concentration ensures optimal growth rate even over several days from plating. EGF is indispensable for MCF10A and CAL148 and has no growth advantages for CAL51 and MCF7.

4.1.15 PCR-Qualitative results

In order to test and verify the efficacy of primer design as well as the effective mRNAs expression of those genes implicated in the γ-secretase complex (APH1A, PSEN2 and NCSTN) and their substrates (CDH1, PYGO2 and BCL9), a PCR was performed (see **Figure 47**).

The APH1A, PSEN2 and NCSTN are differentially expressed genes in the 1q-OverUpT (FCvsCTRL > 1.3 at adjp < 0.05 and FCvsN > 1 at adjp < 0.05) and 16q-UnderT groups (FCvsCTRL< -1.3 at adjp < 0.05 and FCvsN > 1 at adjp < 0.05) (see Pathways analysis par. 4.1.5 – 4.1.6) referred to Group A(der(1;16)). They are significantly deregulated respect to normal samples as well as respect to the "selected CTRL groups not bearing the aneuploidy". APH1A, PSEN2, and NCSTN components are ranked in the top-ten genes of "NOTCH2 Activation and Transmission of Signal to the Nucleus pathways. The PYGO2 and BCL9 are deregulated in 1q-OverUpT and 16q-UnderT groups. These genes are in "Deactivation of the beta-catenin transactivating complex," and "Formation of the beta-catenin:TCF transactivating complex" pathways. CDH1 transcript level is lower in invasive lobular carcinomas either in 16q-loss groups (A, B1, and D) and in the 16q-disomic CTRL-group when compared to corresponding ductal carcinomas or to normal breast tissue (see par. 4.1.8 and 4.1.9 - **Figure 28C**). A previous bioinformatics analysis by using Transcript per million (TPM) values, reveals that those genes are expressed in the studied cell lines (**see Table 20**).

Gene Name	*Chr arm*	*TPM CAL 148*	*TPM CAL 51*	*TPM MCF7*	*TPM MBA-MD-231*
PYGO2	1q21.3	164.7262	73.58466	41.4024386	48.71005341
NCSTN	1q23.2	161.4317	170.4066	44.85264182	142.6508707
PSEN2	1q42.13	49.41786	19.36438	24.15142252	24.3550267
APH1A	1q21.3	**434.8771**	**178.1523**	265.6656477	173.9644765
CDH1	16q22.1	161.4317	46.47452	641.7377984	0
BCL9	1q21.1	98.83572	104.5677	17.25101608	10.43786859

Table 20: Transcript per million (TPM) of the PYGO, NCSTN, PSEN2, APH1A, CDH1, BCL9 in CAL148, CAL51, MCF7, MBA-MD-231.

To verify primer design, APH1A gene was choice as candidate gene in CAL51. APH1A is the most expressed gene in CAL-148 (434.8 TPM) and CAL51 (178.15 TPM), which are cell lines bearing and not bearing the der(1;16), respectively, as reported in **table 20**.

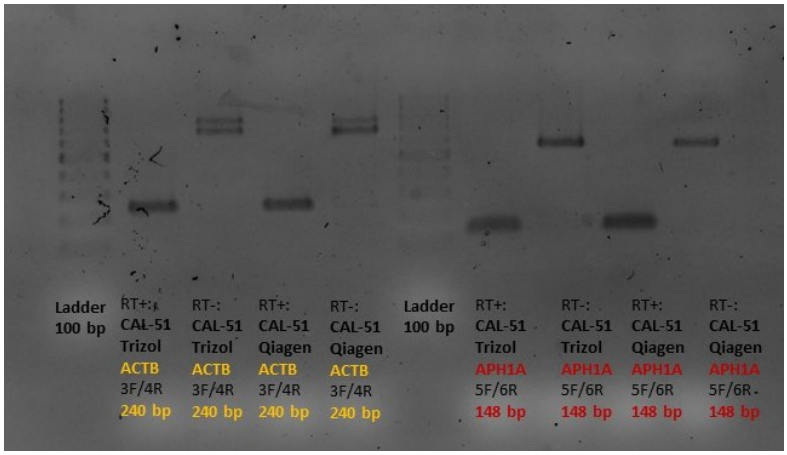


Figure 47: Electrophoretic run in agarose gel

The RNA was extracted by using the Quiagen RNAsy Minikit and an additional method with TRIzol™ Reagent (Cat. No. 15596026 and 15596018; Invitrogen, USA) (see **Table 3**). The cDNA is retrotranscribed from the cell line CAL51 by using Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit™ (cat. n.944404, Applied Biosystems). The internal control was ACTB. The quality checking was performed with RT+ (with primer) and RT- (no primer). As showed in the electrophoretic run in agarose gel in **Figure 47**, the obtained amplicons show the predicted size, confirming the good primer design for APH1A.

The next step, to the previous experiment, was to test the validation of all designed primers (as described in Table 2) in CAL51 cell line. The cDNA is retrotranscribed by Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit™ (cat. n.944404, Applied Biosystems). The electrophoretic run in agarose gel in **Figure 48** shows the obtained amplicons with the predicted size for all genes.

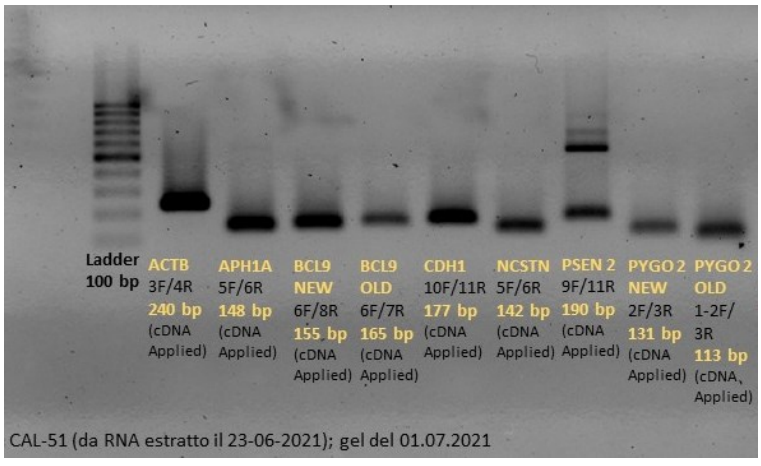


Figure 48: Electrophoretic run in agarose gel:

The **figure 49** shows, in addition to transcript genes in table X, the trend for EFNA1 and EFNA3 (EFNA4 was not present in the data) found to be enriched in 1q-OverUpT and 16q-

UnderT groups and the associated pathways “Ephrin mediated repulsion of cells” with APH1A, PSEN2, and NCSTN.

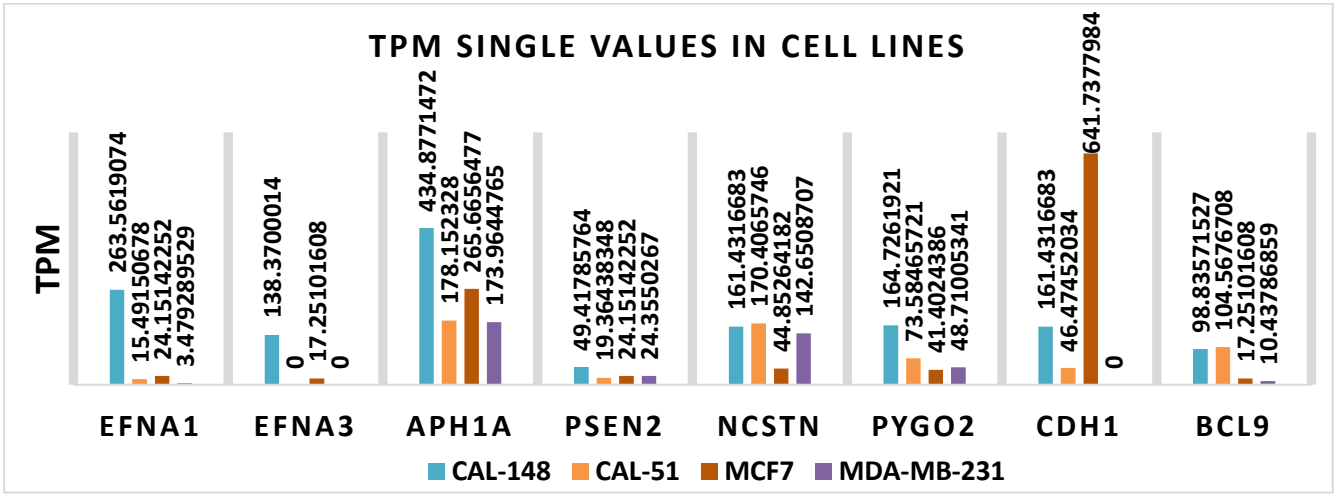


Figure 49: Transcript Per Millions (TPM) for those genes enriched in the pathway analysis for 1q-OverUpT and 16q-UnderT groups

4.1.16 Real-Time PCR: Quantitative results

In order to quantify the mRNA expression of our gene targets concerning APH1A. PSEN2, NCSTN, PYGO2, BCL9 and CDH1, a qPCR reaction using cDNA synthesised from CAL148, CAL51, MCF7 and MDA-MB-231 was performed. The melting curves are showed in **Figure 50** indicating the correctness of the PCR amplicon.

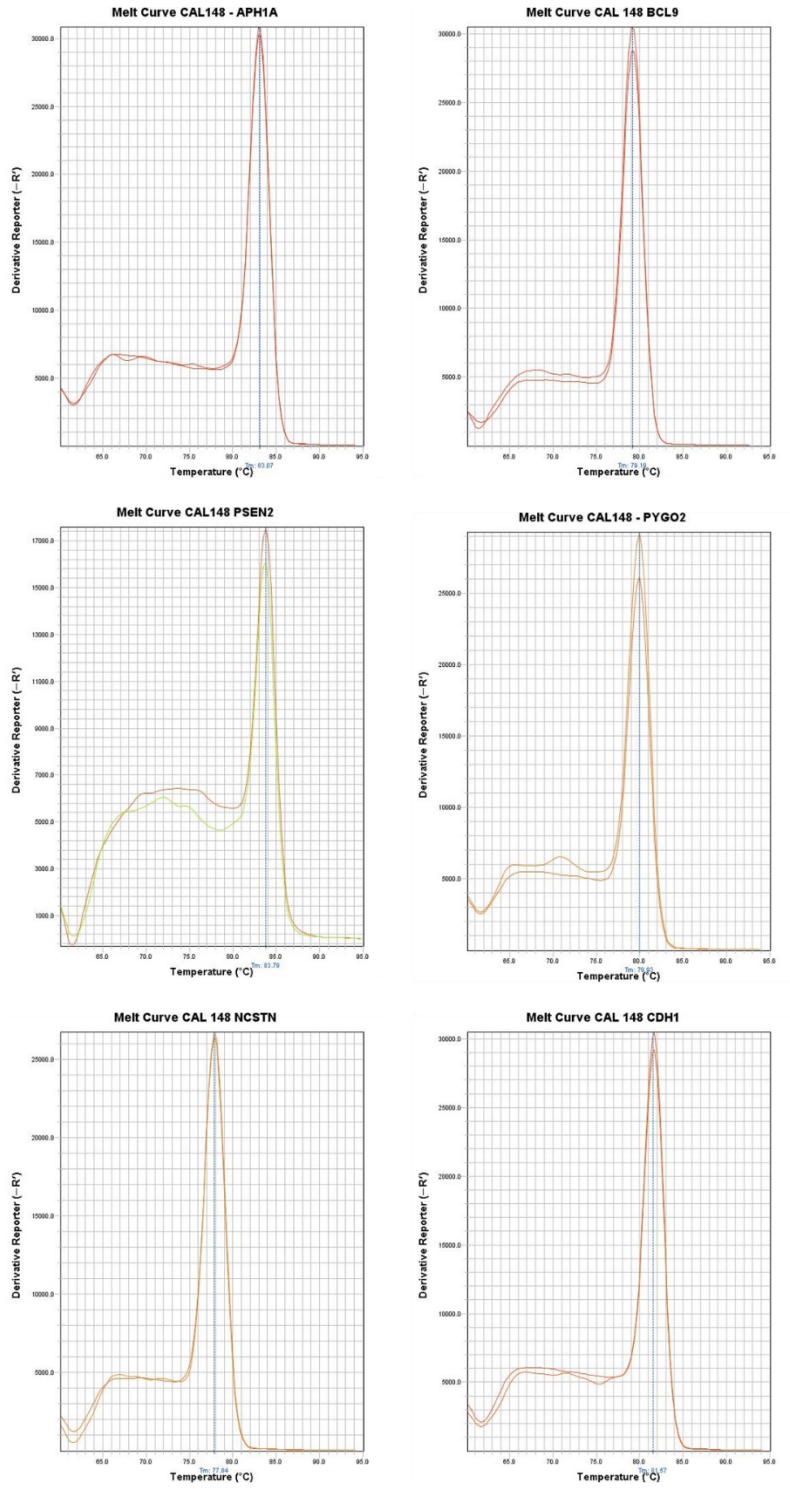
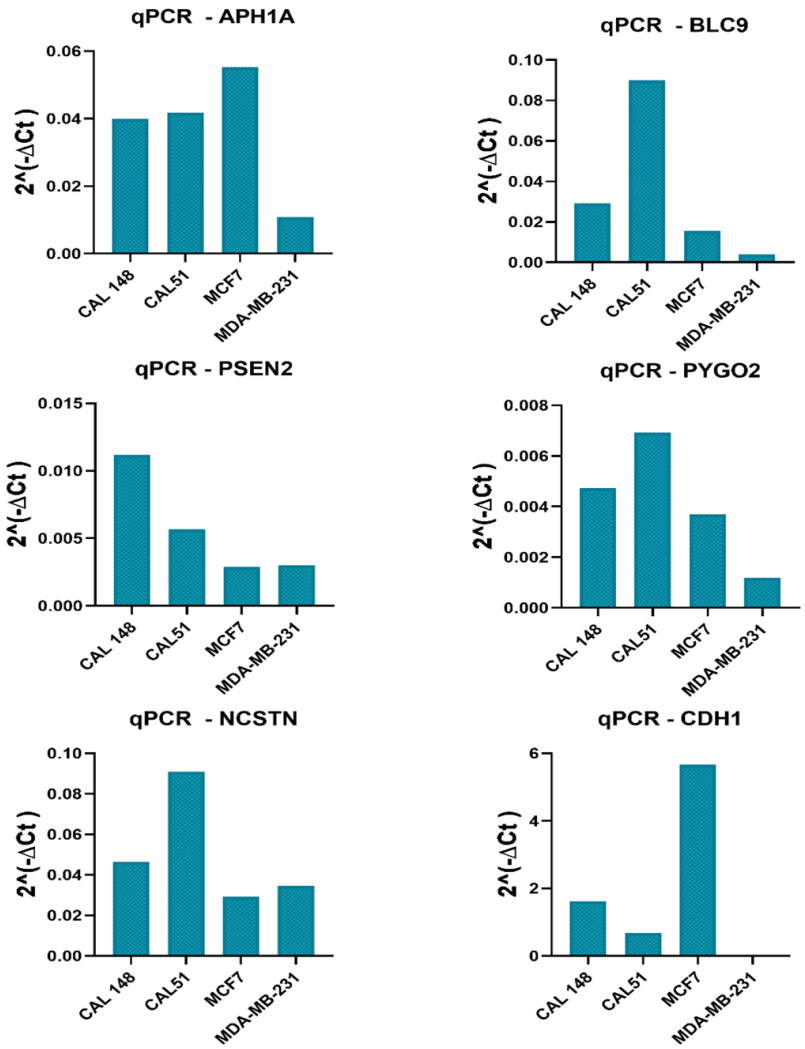


Figure 50. Melting curve analysis of single RNAs in CAL148

For each gene target a threshold cycles (Ct) were produced. The Ct value is the main parameter of real time PCR indicating the number of reaction cycles required so that the fluorescence of each gene exceeds the background fluorescence. Ct values are inversely proportional to the entire amount of mRNA in the sample. For each Ct values a mean for each replicate sample was calculated. Ct were transformed in ΔCt which is a mathematical term to describe the difference between mean of gene target and the mean of housekeeping gene (ACTB). To better interpret this difference(ΔCt) and avoid negative numbers (i.e. if genes are more expressed than housekeeping gene), ΔCt values were converted into linear notation of 2^(-ΔCt).



TPM				
	CAL 148	CAL 51	MCF7	MBA-MD-231
APH1A	434.88	178.15	265.67	173.96
PSEN2	49.42	19.36	24.15	24.36
NCSTN	161.43	170.41	44.85	142.65
BCL9	98.84	104.57	17.25	10.44
PYGO2	164.73	73.58	41.40	48.71
CDH1	161.43	46.47	641.74	0.00

2^-ΔCt				
	CAL 148	CAL51	MCF7	MDA-MB-231
APH1A	0.040049	0.04165	0.055341	0.010795
PSEN2	0.011169	0.005657	0.002879	0.003019
NCSTN	0.046517	0.090756	0.029262	0.034443
BLC9	0.029095	0.089797	0.015613	0.003973
PYGO2	0.004738	0.00692	0.003685	0.00118
CDH1	1.609219	0.680377	5.655101	0.002902

Figure 51. Gene expression pattern of gene target in qPCR

In **figure 51** the relative quantification of each transcript target is observed in CAL148, CAL51, MCF7 and MDA-MB-231. The relative 2^(-ΔCt) (blue table) and TPM (yellow table) are indicated to better perform a comparison between expected and observed results in the cell lines.

The expression trends concerning NCSTN, CDH1 and BCL9 genes confirm the TPM values observed in RNA-seq data in CAL148, CAL51, MCF7 and MDA-MB-231.

APH1A shows important differences in CAL148 by qPCR assay in comparison to TPM. In particular, APH1A TPM expression is higher in CAL148 (434.88 TPM) and MCF7 (265.67 TPM), followed by CAL51 (178.15) and MDA-MB-231 (173.96), instead the relative quantification in qPCR confirm the lower values in CAL51 and MDA-MB-231, but not in CAL148 which shows a lower values and in MCF7 which shows the higher expression values. PYGO2 show minor difference in CAL148 which is lower than expected.

PSEN2 show minor difference in CAL51. In particular, qPCR results confirm TPM values in CAL148, MCF7 and MDA-MB-231 but in CAL51, PSEN2 expression is higher than expected.

To exclude or confirm that the observed differences in relative quantification are due or not to EGF treatment, we have carried out another qPCR experiment evaluating the different effect of EGF treatment on the mRNA expression in CAL148. The results are showed in **figure 52**.

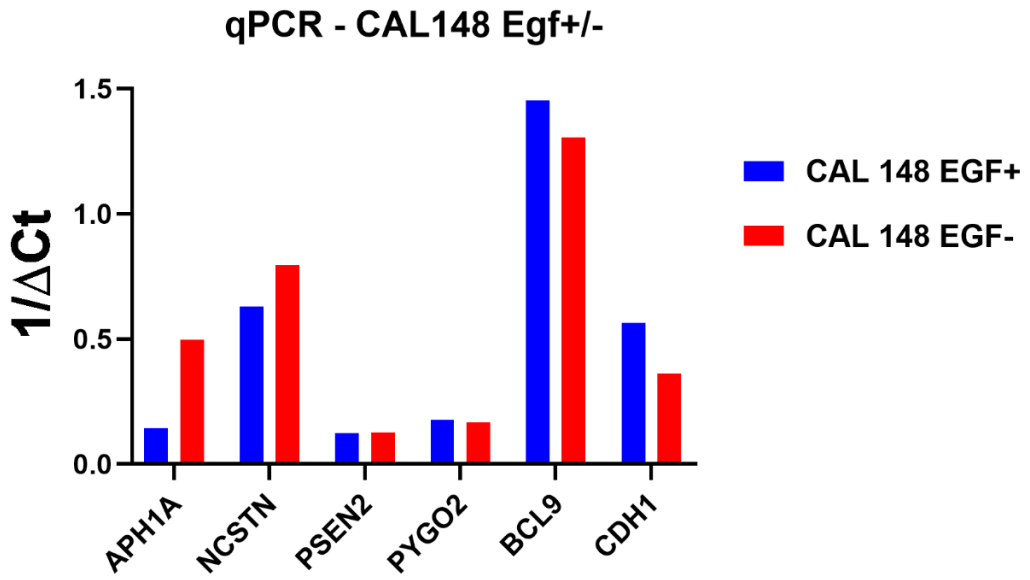


Figure 52. Gene expression pattern of gene target in qPCR in CAL148 Treated and untreated with EGF.

PSEN2 and PYG02 expression do not show any appreciable differences due to EGF treatment in cell population of EGF treated (EGF+) and un-treated (EGF-). BCL9 and CDH1 expressions are influenced by EGF treatment with a slightly higher expression rate.

The most interesting result concerns the trend inversion of APH1A and NCSTN gene expression due to the EGF treatment. If we perform a comparison of qPCR results versus TPM values, APH1A and NCSTN expressions are lower in qPCR results (and then in cell lines) than expected TPM values. Probably, EGF hormone shows a dual effect of activation for some genes and inhibition for others such as APH1A and NCSTN. This dual effect should be re-evaluated in another qPCR experiments and in more effective ways.
While in CAL148 gene expression differences could be due to EGF treatment (i.e. APH1A and NCSTN), in CAL51 they must be due to other factors not fully identified and they will be targets of further future experiments.

4.1.17 Immunocytochemistry (ICC): TBNC testing

Triple-negative breast cancer (TNBC) is one of the most aggressive subtypes of cancer and it is associated with a poor five-year survival rate. It is defined as "triple-negative" because it is characterized by absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2/ERBB2) expression. To classify TBNC, gold standard techniques are the immunohistochemistry (IHC) or immunocytochemistry (ICC) with antibody markers including ER, PR, HER2, CK5/6 and EGFR.
To assess the transcriptomic truthfulness of the well-known and repeatedly discussed data about TBNC in the literature, a bioinformatics checking by analysis of Transcript per Million (TPM) was performed. From an immunohistochemical point of view, CAL-148 cell line have been often classified and used as TBNC (Lang et al., 2020). The transcriptomics data reveals an abundant expression of ERBB2/HER2(200.966 TPM) (see **Table 21**).
In the literature, CAL-148, CAL51, MDA-MB-231 and MCF10A are classified as TBNC (see **table 22**). A study (Subik et al., 2010) reports an incongruent classification example concerning MCF7 which are known to be ER-, but from their IHC re-classification, MCF7 are ER+ (see **Table 23**).

				Lum Her2 Amp	Basal	LumA	Basal
Gene symbol	Chr	band	Description	CAL-148	CAL-51	MCF7	MDA-MB-231
ESR2	14	q23.3	estrogen receptor 2	0	0	0	0
ESR1	6	q25.1	estrogen receptor 1	0	0	55.20325	0
PGR	11	q22.1	progesterone receptor	0	0	10.35061	0
AR	X	q12	androgen receptor	9.883572	0	13.80081	0
ERBB3	12	q13.2	erb-b2 receptor tyrosine kinase 3	161.4317	23.23726	69.00406	0
ERBB4	2	q34	erb-b2 receptor tyrosine kinase 4	6.589048	34.85589	0	0
ERBB2	**17**	**q12**	**erb-b2 receptor tyrosine kinase 2**	**200.966**	**73.58466**	**48.30285**	**24.35503**
EGFR	7	p11.2	epidermal growth factor receptor	13.1781	23.23726	0	104.3787
EGF	4	q25	epidermal growth factor	9.883572	0	0	0
NOTCH1	9	q34.3	Notch receptor 1	23.23726	19.76714	20.87573	13.80081
NOTCH2	1	p12	Notch receptor 2	58.09315	56.00690	90.46152	27.60162
NOTCH3	19	p13.12	Notch receptor 3	50.34739	69.18500	0	72.45426
NOTCH4	6	p21.32	Notch receptor 4	0	0	0	0

Table 21: Transcript per Million (TPM) for those genes refers to the TBNC classification in CAL-148, CAL-51, MCF7 and MDA-MB-231.

	Intrinsic Molecular Subtype	Histological Subtype	IHC	References
CAL-148	Luminal Her2Amp	Ductal	ER- Her2- PR-	www.deepmap.org
CAL-51	Basal		ER- Her2- PR-	www.deepmap.org
MCF7	Luminal A		ER- (ER+) PR+	(Subik et al., 2010)
MDA-MB-231	Basal		ER- Her2- PR-	(Subik et al., 2010)
MCF10A	Basal		ER- Her2- PR-	(Subik et al., 2010)

Table 22: Cell lines subdivision for intrinsic molecular subtype, histological subtype and their IHC status.

	ER	PR	HER2	CK5/6	EGFR	Ki-67	AR	Subtype
MDA-MB-231	0	0	0–1+	–	1+	100%	8	Basal
SKBR-3	0	0	3+	–	2+	20%	8	HER2
MDA-MB-231-UR	0	0	0–1+	–	3+	100%	7	Basal
MCF-12A	0	0	0–1+	+	2+	95%	5	Basal
HBL101	0	0	0–1+	–	1+	90%	4	Basal
MDA-MD-435	0	0	3+	–	0	80%	6	HER2
MCF-7	6	6	0–1+	–	1+	90%	7	Luminal A
HS598T	0	0	0–1+	–	1+	90%	0	Basal
MCF-10A	0	0	0–1+	+	2+	30%	0	Basal
BT-20	0	0	0–1+	–	2+	80%	4	Basal
MCF-10F	0	0	0–1+	+	1+	100%	0	Basal
468	0	0	0	–	3+	95%	8	Basal
AU 565	0	0	3+	–	1+	95%	7	HER2
ZR-75-1	3	4	2+	–	1+	80%	8	Luminal A
BT-483	0	0	0	–	1+	95%	4	Basal
BT-474	0	8	3+	–	1+	70%	7	Luminal B
MDA-MB-453	0	0	0	–	0	80%	8	Unclassfied

Note: For ER, PR and AR, Allred scores were used.; for HER2 and EGFR the scoring system for HER2 was used; for CK5/6, any strong cytoplasmic stain is considered as positive; and for Ki-67, the% of any intensity of nuclear stain was used.

Table 23: The results of IHC analysis for all breast cancer cell lines (Subik et al., 2010)

Hypothesizing a mis-classification of CAL148, CAL-148, CAL-51, MCF10A were tested with ErbB2 (HER-2) Monoclonal Antibody (CB11)(Cat. No. MA1-35720, ThermoFisher) in IHC assay. The test was performed in collaboration with professor Lidia Puzzo and the medical doctor Vecchio Giada from the Department of Medical and Surgical Sciences and Advanced Technologies, G.F. Ingrassia, Azienda Ospedaliero-Universitaria "Policlinico Vittorio Emanuele", Anatomic Pathology, School of Medicine, University of Catania, Italy. The MCF10A were kindly provided by Professor Luca Lanzanó and his collaborator Dr. Morgana D´Amico.

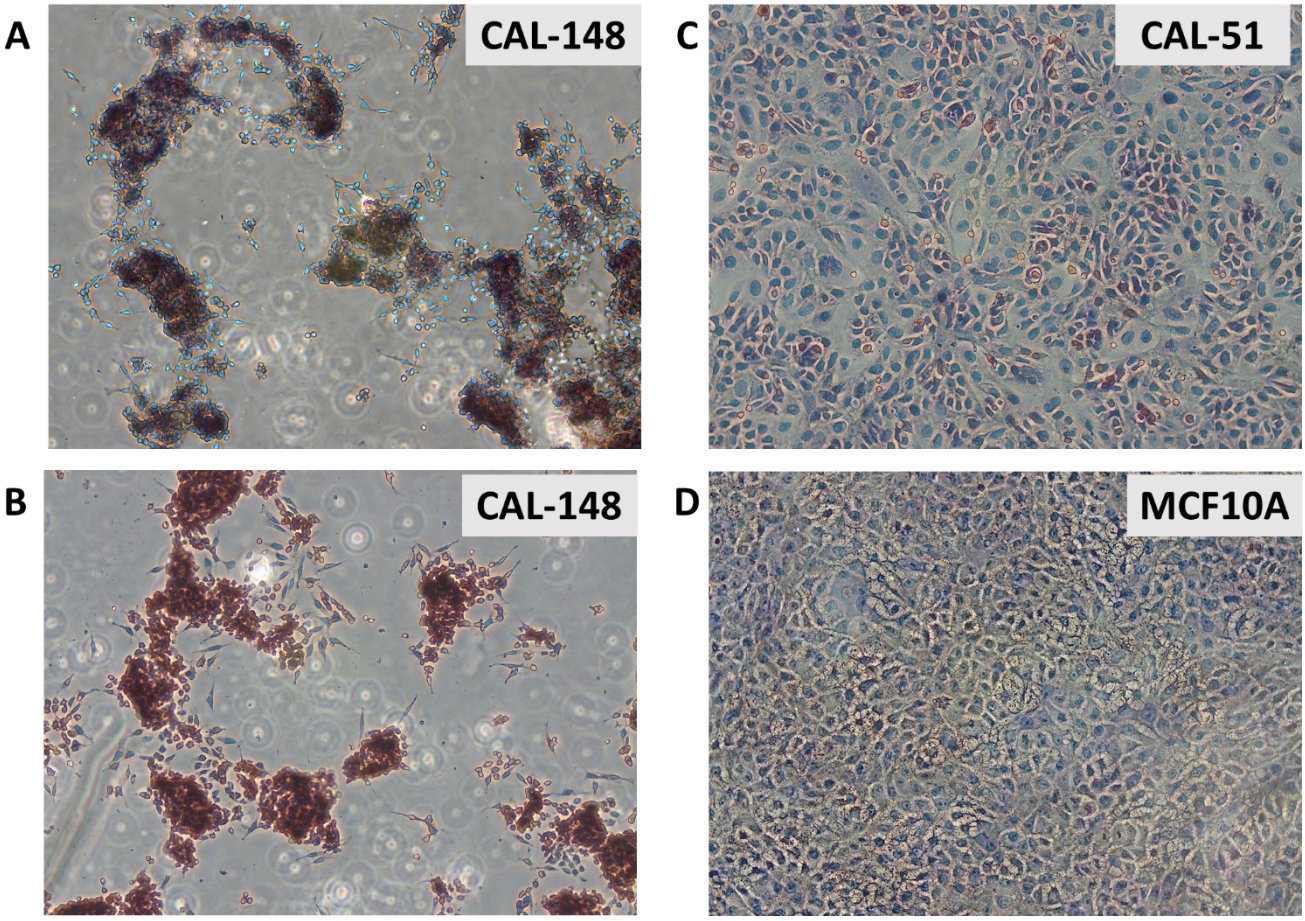


Figure 53 Representative staining results from IHC for HER2 (original magnification 200X). A-B) CAL-148 in brown the positivity to HER2 antibody C) CAL-51 weakly positive to HER2 antibody D) MCF10A negatively to HER2 antibody

CAL-148 are positive to HER2 staining (**Figure 53 A -B**) and CAL-51 are weakly positive to HER2staining (**Figure 53 C**). This result is in contraposition with the literature. (see **table 22**). MCF10A are negative to HER2 staining (**Figure 53 D**).

4.1.18 Cell viability and Gamma secretase inhibitors

In this thesis the effect of γ-secretase inhibitors (GSI), DAPT and PF3084014 (also known as Nirogacestat – commercial name) on CAL148, CAL51 and MCF10A were evaluated. The γ-secretase complex is formed by our gene targets APH1A, NCSTN, PSEN2; one of their main targets is the Notch signalling. In 1,16 chromogroups and more precisely in group A, which is characterized by the aberration der(1;16), the 1q-gain leads to an over expression of the γ-secretase complex components (APH1A, NCSTN and PSEN2) as demonstrated in the previous analysis in **Figure 23**, **Table 20** and **Figure 51**. The over-expression these genes, lead to a hyper-activation of the Notch pathway. The notch signalling is well known to be altered in breast cancer. For the first time the notch alteration is linked to a precise chromosomal aberration and for this reason we have tested the GSI on the cell model, CAL148 bearing the der1;16. We have tested GSI also in the controls cells (CAL51 and MCF10A). The GSI are tested by considering 4 drugs concentrations (for both compounds) of 0.1 μM, 1μM, 10μM, 50μM corresponding to the logarithm of molar concentrations of -7 (0.1μM), -6 (1μM), -5(10μM) and -4.30(50μM). The control group corresponds to the logarithm of molar concentrations of -10 (log of zero is undefined, the first X value to identify the untreated groups is -10 in logarithms notation). The efficacy of the treatment was evaluated with cell vitality assay kit. The absorbance was read by PlateReader AF2200. Wavelength of 600nm. The control group is not explicitly indicated because it was used as normalization group to calculate, for each value, the percentage of inhibition respect to control.

The **figure 54** is a summary of the conducted experiments on CAL148 treated or not treated with EGF at three time-points (24h, 48h, 72h), 4000 Cells/well, CAL51 and MCF10A (only EGF treated at 24h, 48h, 72h). In MCF10A the option without EGF was excluded because the cells are not able to survive without it, while in CAL51 EGF treatment does not give any advantage, but we choose to treat them however in order to uniform experiments and reducing the error variables. A non-linear regression (fit curve) was applied. A dose-response-inhibition curve was produced comparing log(inhibitor) vs. response (three parameters model).

The dose-response curve was also used to determine the half maximal inhibitory concentration (IC_{50}) of the drug. For definition IC50 measures the potency of a substance in inhibiting a specific target or, in this case, the cell vitality. In table 24, IC50 for DAPT and PF3084014 for each cell types and timepoints were calculated.

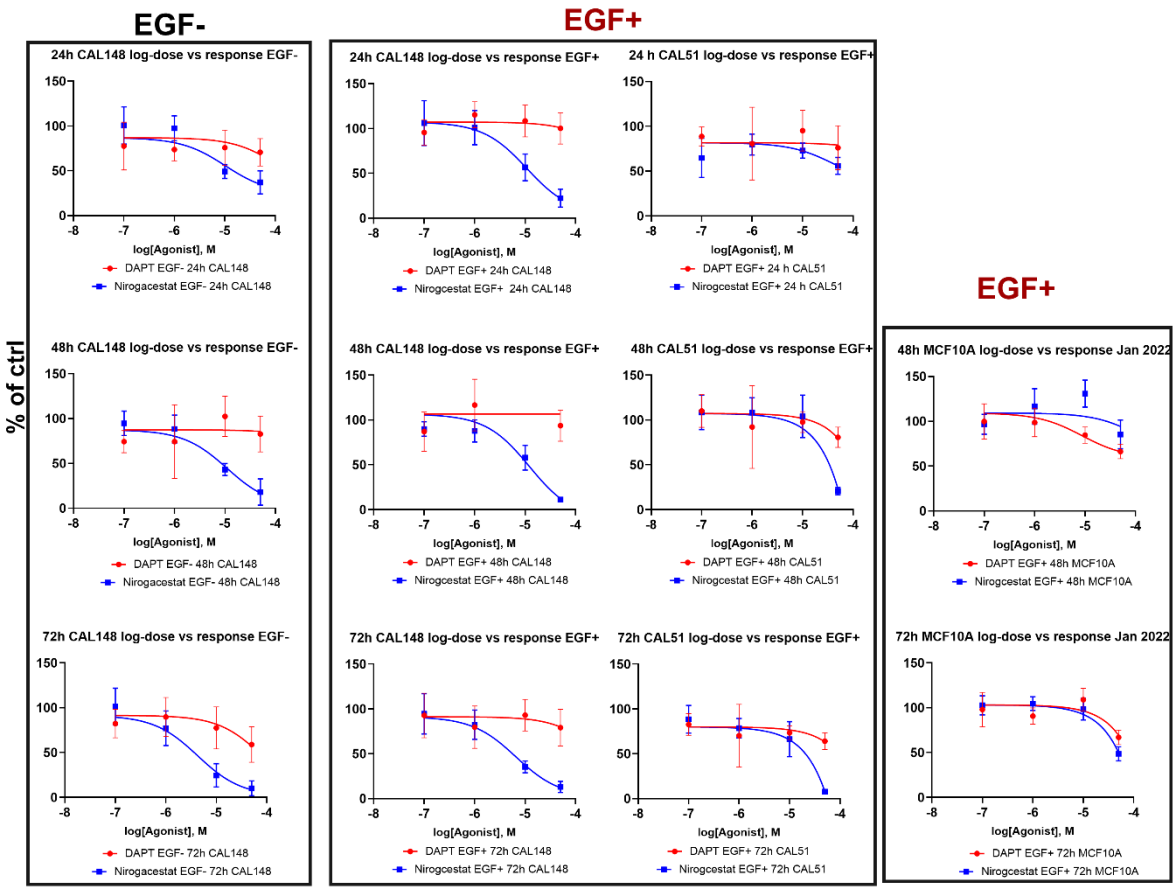


Figure 54: A dose-response-inhibition curves. The X values are the logarithms of molar concentration. The Y values are responses (in replicate) in two experimental conditions CAL 148 - CAL51 - MCF10A- (24h - 48h - 72h) - EGF+-DAPT AND NIROGACESTAT

	24h		48h		72h	
	DAPT	**PF-3084014**	**DAPT**	**PF-3084014**	**DAPT**	**PF-3084014**
CAL-148 EGF- (LogIC50)	**-3.937**	**-4.991**	**-2.636**	**-4.957**	**-4.09**	**-5.369**
IC50 [Agonist], µM	115.7	10.21	2313	11.05	81.28	4.278
95% CI LogIC50	(-4.617 to ??)	(-5.555 to -4.198)	(-3.853 to ??)	(-5.533 to -4.225)	(-4.409 to -3.710)	(-5.709 to -5.033)
95% CI IC50 (µM)	24.14 to ???	2.79 to 63.34	140.40 to ???	2.93 to 59,52	38.95 to 195.10	1.96 to 9.26
Goodness of Fit -R-squared	-0.2089	0.6874	-0.02127	0.8451	0.2159	0.8409
CAL-148 EGF+ (LogIC50)	**-3.117**	**-4.964**	**~ 54605**	**-4.913**	**-3.465**	**-5.181**
IC50 [Agonist], µM	763	10.86	~	12.21	342.9	6.584
95% CI LogIC50	(-3.760 to ??)	(-5.379 to -4.525)	(Very wide)	???	(-3.987 to ???)	(-5.630 to -4.732)
95% CI IC50 (µM)	174 to ???	4.18 to 29.82	(Very wide)	???	103.0 to ???	2.34 to 18.54
Goodness of Fit -R-squared	0.01872	0.81	-0.02305	0.8384	0.01846	0.8554
CAL51 EGF+ (LogIC50)	**-3.054**	**-4.374**	**~ -0.7868**	**~ -1.268**	**~ -1.763**	**~ -2.402**
IC50 [Agonist], µM	882.4	42.28	163400	53960	17250	3965
95% CI LogIC50	(-6.380 to ??)	???	(Very wide)	(Very wide)	(Very wide)	(Very wide)
95% CI IC50 (µM)	0.42 to ???	???	(Very wide)	(Very wide)	(Very wide)	(Very wide)
Goodness of Fit -R-squared	-0.016	0.03914	0.07699	0.8164	0.05042	0.8481
MCF10A EGF+ (LogIC50)	Not Performed		**-5.061**	**-3.923**	**~ -0.9865**	**~ -1.183**
IC50 [Agonist], µM			8.683	119.3	103200	65560
95% CI LogIC50			(-6.532 to ??)	(-4.539 to ??)	(Very wide)	(Very wide)
95% CI IC50 (µM)			0.29 to ???	28.91 to ???	(Very wide)	(Very wide)
Goodness of Fit -R-squared			0.4489	0.1387	0.4088	0.8537

Table 24: Half maximal inhibitory concentration (IC_{50}) in logarithms notation and in red the µM IC50 concentration. The 95% confidence interval and the Goodness of Fit -R-squared.

By analysing the obtained results (**Figure 54** and **Table 24**) the main result concerns the effect of PF3084014 on CAL148 EGF-treated (EGF+) and CAL148 EGF-untreated (EGF-). The DAPT effect is completely absent or active at high concentration, and the curves show a flat trend in the graphs. Ran et al have demonstrated the different pharmacologic effect of GSIs showing that DAPT is the less active on NOTCH receptors in comparison to PF3084014 and other GSIs (Ran et al., 2017). For these reasons DAPT can be considered as a poorly-active control drug in any performed experiments. In particular, we observe a drastic cell viability decline, already at 24h and 10μM (-4 Log[μM]), in CAL148 treated with PF3084014. There is a slightly IC50 increasing concentration in CAL148 EGF+ (4.278 μM at 24h; 11.05μM at 48h, 6.584 at 72h) in comparison to CAL148 EGF- (10.21μM at 24h; 12.21μM at 48h; 4.278μM at 72h). In CAL51, PF3084014 is slightly active at 42 μM (at 24h), while is inactive at 48h and 72h. In MCF10A potencies of DAPT and PF3084014 are ambiguous at 48h and inactive at 72h. These experiments give us an indication of the inhibiting effect of GSIs on our cell lines and especially on CAL148.

4.1.19 PSEN2 siRNA interfering treatment

In order to further analyse the role of the gamma secretase complex in breast cancer, a siRNA interfering treatment protocol has been set up. The silencing was performed on MCF10A (non-cancerous cell) by using the siRNA PSEN2 (Cat.No. 1027416, 1nmol siRNA, FlexiTube Gene Solution, Qiagen). In **Figure 55A** the transfection was conducted at 48h and 72h, at three siRNA concentration (10nM, 20nM, 40nM). The un-treated cells group (CTRL MCF10A) and the non-targeted siRNAs group were evaluated. In **Figure 55B** the transfection was repeated at 42 and 72h on CAL148 by using siRNA PSEN2 at 40nM.

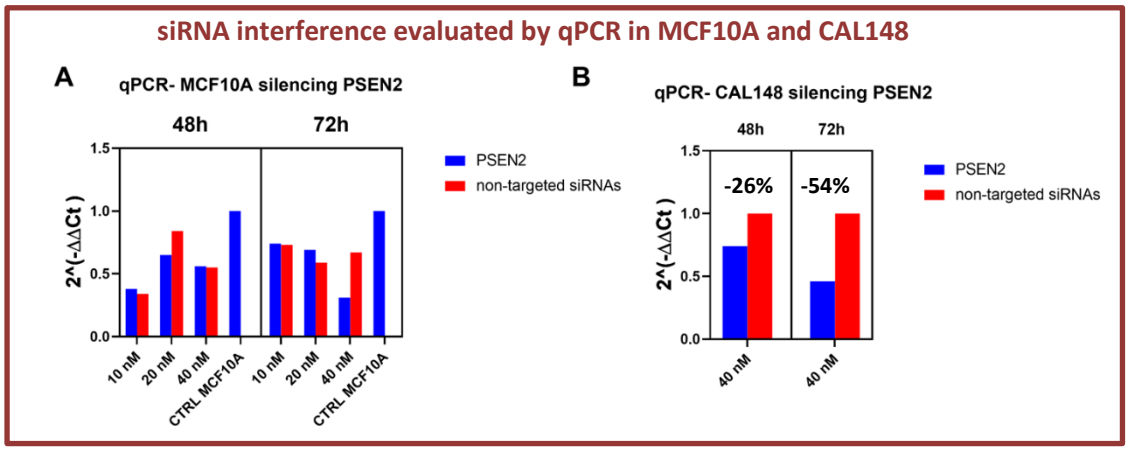
The siRNA efficacy was evaluated in qPCR and the obtained results are then converted by the 2^(-ΔΔCt) method by normalizing the unit to the housekeeping gene (ACTB). The **figure 55A-B** show the results.

In **Figure 55A** The MCF10A was successfully transfected. At 48h there is a gene reduction at 20nM in treated group (blue histograms) in comparison to non-targeted siRNAs group (red histograms). At 72h the effect at 20nM concentration disappeared, maybe due to the high cell growth rate of MCF10A, while we observed a high reduction of treated groups in comparison to the non-targeted siRNAs group. However, both treated groups and un-treated groups show lower values in comparison to MCF10A CTRL.

In **Figure 55B,** we tested the transfection on CAL148 at 48h and 72h considering a siRNA concentration of 40nM (the most effective concentration in MCF10A). The data were normalized against the non-targeted siRNAs values (non-targeted siRNAs value is used as calibrator in the 2^(-ΔΔCt) method) The mRNA expression is reduced of around 26 % at 48h and 54% at 72h.

Data in **Figure 55 C-G** concern results of siRNA interference coming from several experiments and evaluated with MTT cell viability assays. Data were normalized by removing the background and then the values were expressed as percentage of the non-targeted siRNA group. The statistical significance was evaluated with un-paired t-test.

In **Figure 55 C-D** the siRNA efficacy was tested on CAL148 by considering six biological replicates. Results concern treatments with siRNAs of APH1A, NCSTN and PSEN2 at 40nM in comparison with non-targeted siRNAs cell group (**Figure 55C**). siRNAs were also tested in combination (APH1A 40nM + PSEN2 40nM) and (NCSTN 40nM + PSEN2 40nM) (**Figure 55D**).



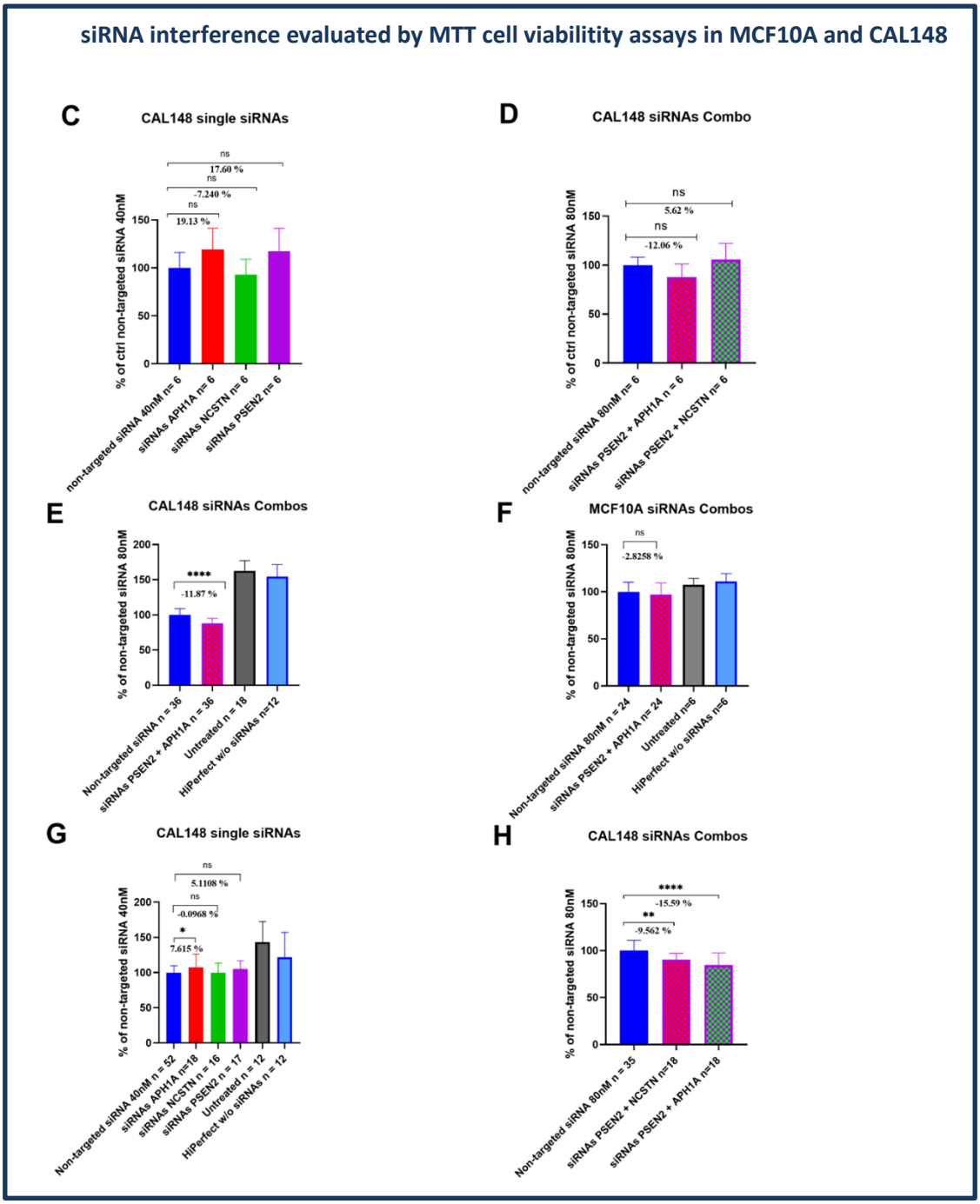


Figure 55: A) qPCR evaluation on MCF10A at 4000 Cells/well concentration and treated with PSEN2 siRNAs at three siRNA concentration (10nM, 20nM, 40nM). Two Time points (48h, 72h) were considered

B) qPCR evaluation on CAL148, siRNA PSEN2 at 40nM concentration. Two time points (48h -72h) C-D) MTT cell viability assay to evaluate siRNA interfering (singles siRNAs and Combo PSEN2 + APH1A and PSEN2 + NCSTN) on CAL148 (6 replicates) E) MTT cell viability assay to evaluate siRNA interfering (Combo PSEN2 + APH1A) on CAL148 by considering 36 replicates F-G) single siRNAs interference on CAL148 (F-H) and on MCF10A (G) by considering 12-52 replicates.

In **Figure 55 C-D** the combination of two siRNAs (APH1A and PSEN2) shows a reduction of 12.06 % in cell viability in comparison to the non-targeted siRNA group at 80nM (**Figure 55D**) while single siRNA NCSTN silencing (**Figure 55 C**) shows only a reduction of 7.240 % but both results are not statistically significant. These preliminary results are due to an exiguous number of biological replicates. In order to re-test the obtained results and made them reproducible, we re-test the siRNA Combo of PSEN2 + APH1A in comparison to non-targeted siRNA group. The number of biological replicates was drastically improved at thirty-six. The biological significance was tested with un-paired t-test (**Figure 55 E**). The siRNAs Combo PSEN2 + APH1A produce a reduction of 11.87 % cell viability in CAL148. This result is confirmed by no significant reduction (-2.82% cell viability) in MCF10A (control cell line) (**Figure 55F**). Regarding the transfection of CAL148 with single siRNAs (APH1A, NCSTN, PSEN2), results in **Figure 55G** indicate that single siRNAs treatments do no reduce cell viability in CAL158. On the contrary, we observed an increase oef cell viability in those groups of CAL148 treated with single siRNAs APH1A (+7.61% of cell viability in comparison to the non-targeted siRNAs cell group) (**Figure 55G**).

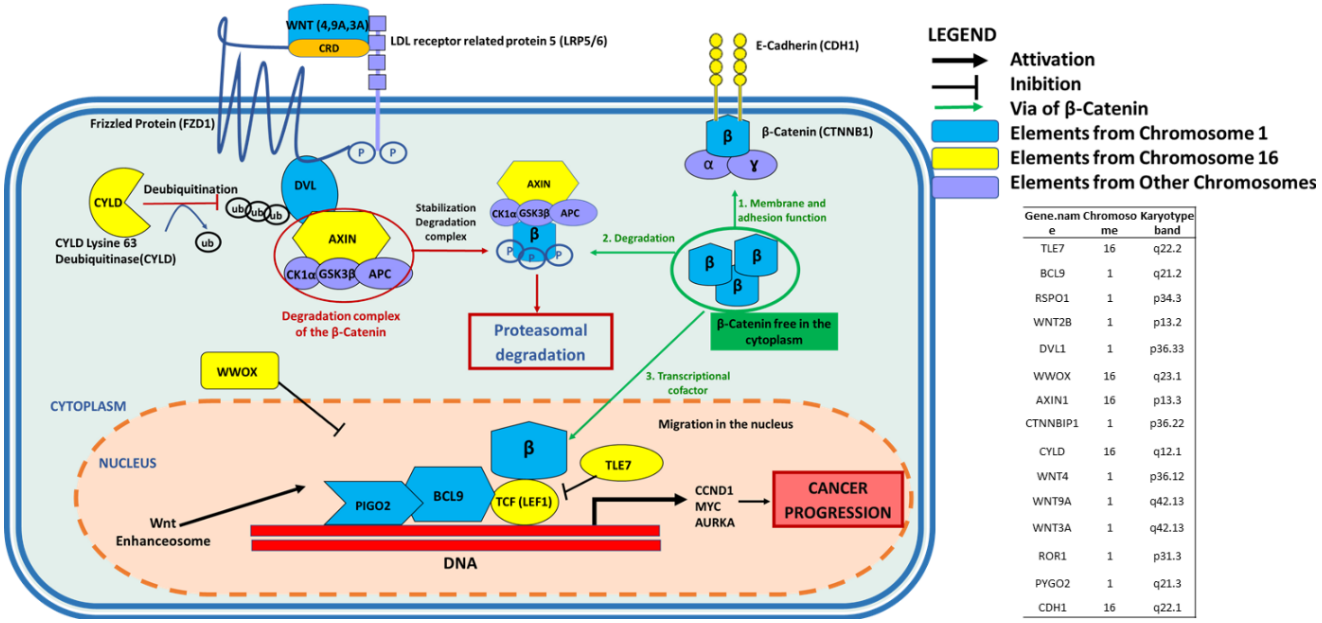
The most important results are indicated in **Figure 55H**. In CAL148, siRNAs transfection performed in combo produces a significant reduction of cell viability. In particular, siRNAs combo PSEN2 + NCSTN treatment reduces of -9.52% cell viability while siRNAs combo PSEN2 + APH1A treatment of -15.59% (**Figure 55H**).

These pilot experiments were a first approach to the siRNA interfering. The future efforts should be directed to the improvement of the transfection procedure in our cell model CAL148 and CAL51 as well as in MCF7 and MDA-MB-231. The parameters to adjust involve, certainly, the synchronous use of all three siRNAs of the gamma secretase complex (APH1A, PSEN2 and NCSTN) only or in combination of another siRNA target (such as BCL9). The experiments were performed with the precious support of my phD colleagues Chiara Scuderi and Virginia di Bella from the Department of Biomedical and Biotechnological Sciences, Section of Medical Biochemistry, University of Catania, Italy.

4.1.20 Pathways model: Wnt enhanceosome and beta-catenin formation

Among 1,16-chromogroups substantial differences in terms of other chromosomal aberrations, aneuploidy scores, transcriptomic data, and single-point mutation profiles were identified. Such information, integrated with a comparative pathway analysis among different 1,16-chromogroups, suggest novel functional links among transcriptionally dysregulated genes in 1q and 16q in invasive ductal and lobular breast carcinomas.

In **Figure 56**, a pathways model concerning the Wnt enhanceosome and the beta-catenin formation is proposed to explain one of possible cooperation mechanisms between chromosome1q-gain and 16q-loss. As indicated in the legend in **figure 56**, the blue colour elements identified the transcripts derived from chromosome 1, while in yellow the transcripts derived from chromosome 16.



Gene.name	Chromosome	Karyotype band
TLE7	16	q22.2
BCL9	1	q21.2
RSPO1	1	p34.3
WNT2B	1	p13.2
DVL1	1	p36.33
WWOX	16	q23.1
AXIN1	16	p13.3
CTNNBIP1	1	p36.22
CYLD	16	q12.1
WNT4	1	p36.12
WNT9A	1	q42.13
WNT3A	1	q42.13
ROR1	1	p31.3
PYGO2	1	q21.3
CDH1	16	q22.1

Figure 56: Pathways design about the proposed cooperation mechanism

Among the genes located on chromosome 16q, the CDH1 gene, encoding for the cell adhesion glycoprotein E-cadherin, has been repeatedly implicated as an important player in mediating the effect of 16q-loss in breast cancer. Indeed, CDH1-inactivating mutations have been found in 15–56% of invasive lobular breast carcinomas, and the majority of such mutations are associated with 16q-loss, thus generating the typical biallelic inactivation of tumour suppressor genes (Berx et al., 1996; Ciriello et al., 2015; Desmedt et al., 2016; Droufakou et al., 2001; Huiping et al., 1999). On the contrary, invasive ductal breast carcinomas rarely harbor CDH1-inactivating mutations (Berx et al., 1996; Huiping et al., 1999). However, both histological subtypes have shown a frequent 16q-loss, independently by the presence of inactivating point mutations of CDH1, and the invasive lobular carcinomas

have shown a reduced expression of CDH1 at both the mRNA and protein level (Alsaleem et al., 2019; Berx et al., 1996; Ciriello et al., 2015; Gamallo et al., 1996; Michaut et al., 2016).. It is well-known that E-cadherin antigen, detected by immunohistochemistry analysis, is mainly expressed in ductal carcinomas and absent in lobular ones (Acs et al., 2001; Choi et al., 2008; Dabbs et al., 2013; Grabenstetter et al., 2020). In agreement with immunohistochemistry results, a previous analysis of TCGA data (Ciriello et al., 2015) reported that CDH1 transcript and protein levels are significantly lower in lobular carcinomas compared to ductal ones, confirming that CDH1 expression differentiates the two histological subtypes. In the present work, we reported that the CDH1 differential expression between ductal and lobular carcinomas was maintained in all examined 1,16 chromogroups that included a significant number of both histotypes (**Figure 28**).

E-cadherin(CDH1) is one of the substrates of gamma-secretase (Marambaud et al., 2002). Gamma-secretase cleavage dissociates E-cadherin from the cytoskeleton, thus promoting the disassembly of the adhesion complex and increasing the cytosolic pool of beta-catenin (CTNNB1). As shown in the **figure 56** the free beta-catenin in the cytoplasm can come up against three different outcomes (see green circle and arrows):

1) **re-building membrane adhesion complex:** β-catenin is a component of adherents junctions and its function is creating a link between E-cadherin and the actin cytoskeleton (Clevers et al., 2014)
2) **to be eliminated by proteasomal degradation complex:** the β-catenin is continuously phosphorylated by the degradation complex that includes the proteins AXIN, APC, GSK3β and CK1α, which marks the protein by directing it to proteasomal degradation (Larion et al 2017).
3) **to migrate into nucleus where can acts as transcriptional cofactor:** the increase of cytosolic beta-catenin, due to adhesion complex disassembly, allows for its translocation to the nucleus where beta-catenin plays a crucial role in the so-called "Wnt enhanceosome." Such a multiprotein complex, containing beta-catenin, BCL9, Pygo, and TCFs (T cell factors), activates the transcriptional program of Wnt signaling (Mieszczanek et al., 2019). BCL9 functions as a scaffold of the Wnt enhanceosome by binding to the Pygo protein and to the N-terminus of the armadillo repeat domain of β-catenin, as well as by stabilizing the interactions of beta-catenin with TCF bound to cis-regulatory enhancers of Wnt-responsive genes (van Tienen et al., 2017).

Elsarraj et al. have showed a role for BCL9 in the transition from in situ to invasive ductal breast carcinoma and reported that BCL9 knockdown is able to inhibit the proliferation,

migration, and invasion of ductal carcinoma (Elsarraj et al., 2015). Moreover, they also analysed TCGA gene expression data reporting that the Wnt/β-catenin pathway is significantly upregulated in BCL9-high cancers compared to BCL9-low breast cancers. In agreement with those data, our present analysis of OverUpT 1q genes and UnderT 16q genes pointed out the possible functional relevance of the pathway "Formation of the beta-catenin:TCF transactivating complex." In particular, both BCL9 and its interacting partner, PYGO2, are located on chr1q and are overexpressed and upregulated in 1q-gain cytogenetic groups. Interestingly, BCL9 belongs to a small subset of 1q genes that were also found to be overexpressed and upregulated in group D, a cancer group bearing 16q-loss and 1q-disomy. Indeed, other mechanism linking BCL9 to Wnt signaling have been described, such as the ability to inhibit clathrin-mediated degradation of LRP6 signalosome components (Chen et al., 2018).

The hypothesis of a central role of BCL9 and Wnt signaling pathway may have distinct functional implications in the pathogenesis of ductal and lobular breast cancers. As previously discussed, the profound downregulation of E-cadherin is considered a hallmark of lobular cancer cells, and it has been reported that the loss E-cadherin in lobular cancers is associated with the destabilization of the beta-catenin protein, resulting in impaired canonical Wnt signaling (Borcherding et al., 2018; De Leeuw et al., 1997; Shackleford et al., 2020). On the other hand, a decreased functionality of E-cadherin can be achieved in ductal carcinomas by an excessive gamma-secretase processing (as hypothesized above), but this kind of mechanism can partially preserve E-cadherin membrane expression. Several research papers have reported an increased nuclear beta-catenin accumulation and an increased activity of beta-catenin dependent transcriptional activity in breast cancer (Lin et al., 2000; Mukherjee et al., 2012, 2016; Mukherjee & Panda, 2020; Tentler et al., 2020; M. Wang et al., 2015; X. Xu et al., 2020; Y. Xu et al., 2020). However, the nuclear accumulation of beta-catenin or increased beta-catenin dependent transcription have only been detected in subgroups of breast cancers identified as triple-negative ones. Since our analysis revealed an overexpression of BCL9 in estrogen-receptor positive Lum A cancers, it is possible that BCL9 interacts with proteins other than β-catenin, and its activity may be, in part, independent of Wnt/β-catenin (Jiang et al., 2020). In this regard, it is interesting that BCL9 binds to proteins that transmit signals from estrogen receptor, thus connecting its overexpression to estrogen receptor-dependent transcriptional activity (van Tienen et al., 2017). Moreover, in the case of invasive lobular carcinoma, it has been reported that a member of the Wnt protein family, WNT4, is transcriptionally induced by estrogen receptors and drives non-canonical Wnt signaling in

lobular cancer cells (Shackleford et al., 2020). Therefore, although pathway analysis connected BCL9 and PYGO2 protein to the "beta-catenin:TCF transactivating complex," it is possible that those proteins play a special role in beta-catenin-independent signaling pathway both in ductal and lobular carcinomas.

The WWOX gene is located at 16q23.1–23.2, in a region containing the common fragile sites FRA16D, and its deletions have been observed in a large number of breast cancer cases (Pospiech et al., 2018). In epithelial cells, WWOX, a WW-domain containing protein, modulates gene transcription through interaction with p73, AP-2gamma, and ERBB4 proteins (Pospiech et al., 2018; Schuchardt et al., 2013). Pathway analysis by the Reactome database pointed out the significant involvement of the WWOX gene, only in subgroup D2, in the pathway "Nuclear Signaling by ERBB4." Such a pathway can be considered another example of the interaction between the 16q gene WWOX and the gamma-secretase complex subunits encoded on chr1q, since WWOX binds to a cytosolic fragment of the membrane receptor ERBB4, which is generated by the gamma secretase complex, and prevents its translocation to the nucleus (Aqeilan et al., 2005). Moreover, WWOX has been reported as an inhibitor of the Wnt pathways (Bouteille et al., 2009) by its interactions with the three members of the Dishevelled (Dvl) family. Therefore, the decreased expression of WWOX in 16q-loss cancers can contribute to the hyper-activation of Wnt pathways.

The use of curated knowledge-bases, such as Reactome, allows one to explore gene interactions with a certain degree of confidence in the experimental validation of functional pathways. However, the identification of novel interactions or the dissection of complex interactions can be difficult to attain by this methodology. Several interactions reported in the scientific literature are not necessarily revealed by this type of analysis. A first example is represented by the deubiquitinating enzyme CYLD, the familial cylindromatosis tumour suppressor gene, that acts as a negative regulator of proximal events in Wnt signaling at the level of the Dvl proteins, thus potentially having a role in both beta-catenin-dependent and -independent Wnt pathways (Tauriello et al., 2010). Indeed, CYLD is located in chromosome 16q and its expression is reduced in 16q-loss cancer groups (A, B1, and D), thus suggesting that its decreased function could cooperate in the hyperactivation of conventional or non-conventional Wnt pathways.

4.1.21 Notch signalling and gamma secretase complex

NOTCH signalling is a cell-to-cell communication system composed by transmembrane Notch receptors (Notch1–4) and transmembrane ligands (Delta/Jagged). After ligand binding, Notch receptors undergo conformational changes that expose a proteolytic site in the extracellular region. The involvement of Notch signalling system in breast cancer has been repeatedly suggested in the literature (Mollen et al., 2018). After cleavage, the remaining membrane fragment is cleaved at an intramembrane (inner leaflet) site by the gamma secretase complex, thus generating a soluble Notch-intracellular domain (NICD) that is able to translocate to the nucleus activating a specific transcriptional program. Some subunits of the gamma-secretase complex were found to be encoded by genes located on 1q (APH1A, PSEN2, and NCSTN) and were overexpressed and upregulated in cytogenetic groups A, B, and C (1q-gain bearing groups). Indeed, it has been previously shown that NCSTN (nicastrin) is overexpressed in breast cancer, and its genetic depletion is sufficient to inhibit tumour growth in vitro and in vivo (Filipović et al., 2011). The increased NCSTN copy-number, due to 1q-gain, can enhance other transcriptional and post-transcriptional mechanisms, thus leading to hyper-activation of gamma-secretase and NOTCH signaling in breast cancer (Villa et al., 2014).

The results of pathway analysis revealed an interesting functional link between gamma secretase genes, located on 1q, and genes located on 16q. WWP2 is a 16q gene that encodes an E3 ubiquitin-protein ligase acting on Notch3-NICD and targeting it to an endosomal/lysosomal degradation fate (Jung et al., 2014). Indeed, the decreased WWP2 expression, associated with 16q-loss, can contribute to the pathological hyperactivation of Notch3-dependent gene expression. A role of Notch3 hyperactivation has been also shown in experimental models of breast ductal cancerogenesis (Hu et al., 2006; Ling et al., 2013; Pradeep et al., 2012).

Another example is provided by the CBFB gene (Core-Binding Factor Subunit Beta), the beta subunit of a heterodimeric core-binding transcription factor that has been reported as frequently mutated in breast cancer (Banerji et al., 2012); it is located in chromosome 16q and shows a decreased expression in 16q-loss chromogroups (**Figure 23B**) and a high frequency of point mutations in group A and B1. Indeed, a role for CBFB in the suppression of breast cancer has recently emerged, and it has been reported that nuclear CBFB/RUNX1 complex represses the oncogenic NOTCH signaling pathway in breast cancer (Malik et al., 2019). Moreover, an efficient function of the CBFB/RUNX1 complex is necessary for the maintenance of the normal mammary epithelial phenotype (Rose et al., 2020). In particular,

the CBFB/RUNX1 complex represses NOTCH3 (Malik et al., 2019), and this observation establishes another interesting link among the underexpression of a 16q gene (CBFB)—the overexpression of 1q-located gamma-secretase component and the pathway "NOTCH3 Activation and Transmission of Signal to the Nucleus," as described in previous paragraphs.

4.2 COLON- ADENOCARCINOMA (COAD)

In the present work of thesis, we extend and improved method of previous work (D. F. Condorelli et al., 2018) and the model of approach (as previously discussed in par. 1.4 Experimental design to investigate aneuploidies). The data analysis was performed on colon cancer samples provided by The Cancer Genome Atlas (TCGA) consortium, showing that the chromosomal density of cancer up-regulated genes in selected aberrant chromosomes is correlated to the frequencies of broad copy number gains (BCNGs) in the same chromosomes. Moreover, we analyse the effects on protein-coding and non-coding transcripts as well as evaluate the possible correlation between aberrant gene enhancer hyperactivity and BCNG-associated transcriptional effects. Finally, we took advantage of the availability of studies by high-resolution CRISPR screens that have defined a cancer fitness gene as any gene whose knockdown decreases cell growth and proliferation in cancer cell lines (Hart et al., 2015), and sought to establish the relationship between cancer fitness properties and BCNG transcriptional dysregulation.

4.2.1 Chromosomal Distribution of Arm-Level Copy Number Abnormalities

Among Copy Number Abnormalities (CNAs), we selected arm-level CNAs defined as somatic chromosomal aberrations that involve more than 50% of a chromosomal arm (p or q). In **Figure 57A** we report the chromosomal distribution of arm-level gains and losses in the examined population of 433 colon adenocarcinoma (COAD) samples (data obtained by Affiymetrix SNP 6.0 array and downloaded from GDC Data Portal, NIH, National Cancer Institute, https://portal.gdc.cancer.gov/).

In agreement with previous data in colorectal cancer (Barresi et al., 2017) a high frequency (>20% of the analysed samples) of arm-level gains can be observed in chromosomes 20p/q, 13, 7p/q, 8q, and a high frequency of arm-level losses in 18p/q, 8p, 17p, 14, 15, 4p/q, 21, 22, 1p (**Figure 57A**).

In several cases p-arm loss of a chromosome, such as chromosome 8 or 20, was associated with a q-arm gain of the same chromosome. This kind of alteration is remindful of a type of cytogenetic abnormality, called “isochromosome”, which is characterized from the loss of a chromosomal arm accompanied by the duplication of the remaining one, and is abbreviated as i(chromosome number and duplicated arm). Indeed, isochromosomes have frequently been cytogenetically detected in human cancer (Barbouti et al., 2004). Although data on copy number changes do not formally prove the presence of an isochromosome, in the present

paper we indicate as “isochromosome” any simultaneous loss of a chromosome arm associated with the gain of the remaining one. According to such an understanding, we calculate that i(8q) is present in 24.2% of COAD samples, i(20q) in 13.4%, i(17q) in 8.7%, i(1q) in 6.7%, i(5p) in 4.6%.

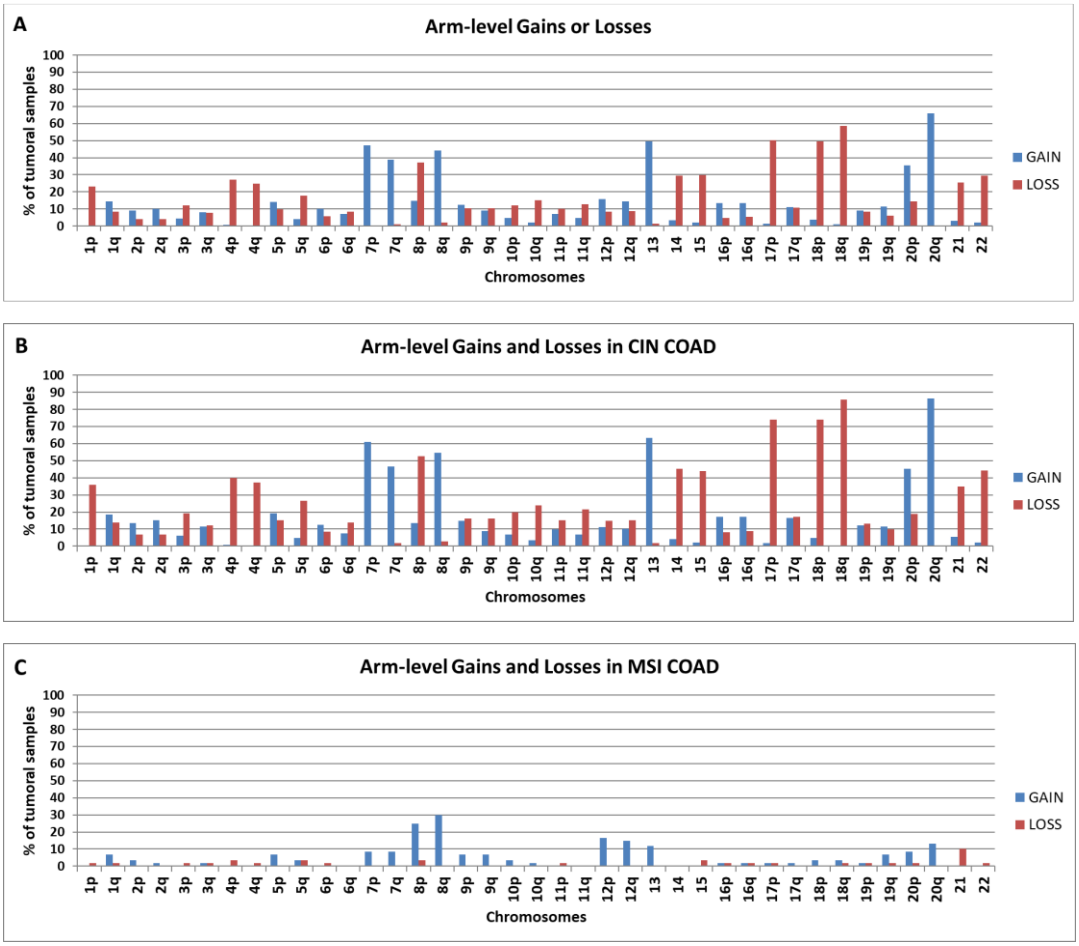


Figure 57. Results are expressed as the percentage of tumour samples bearing arm-level gains or losses in the p or q arm in all colon adenocarcinoma (COAD) samples (A), in Chromosomal INstability colon adenocarcinoma (CIN COAD) samples (B) or MicroSatellite Instability (MSI) COAD samples (C).

Moreover, we separately analysed 223 samples showing Chromosomal INstability (CIN) and 60 samples showing MicroSatellite Instability (MSI), using the clinical information provided by cBioPortal (www.cbioportal.org) (Cerami et al., 2012; Gao et al., 2013)] according to the definitions offered by Liu et al.(Y. Liu et al., 2018). Results on CIN samples (**Figure 57B**) are similar to those observed in the entire group of COAD samples with an obvious increase of the frequencies due to the higher number of arm-level CNAs per sample. Quantitative and qualitative differences can be observed in MSI samples (**Figure 57C**) in comparison to CIN ones.

The decreased frequency of arm-level CNAs is in agreement with the well-known observation that the majority of MSI colorectal samples show a nearly normal karyotype (Barresi et al., 2017). Indeed, in the present study, 26 MSI tumours out of 60 (43.33%) do not bear any broad CNAs, and are considered as normal karyotype tumours. Moreover, other 15 MSI samples bear less than or equal to two arm-level CNAs, and are considered as a near-normal karyotype. In total, 68% of MSI tumours show a normal or a near-normal karyotype. In contrast, only 2% of the 223 CIN tumours show less than or equal to two broad CNAs. In MSI samples, arm-level gains in 8p/q are the most frequent chromosomal CNAs, while 18q losses, one of the most frequent aberrations in CIN COAD samples, are almost absent.

4.2.2 Differential Expression Analysis of RNA-Seq Data in TCGA COAD Samples

Analysis of differential expression of transcript levels between all COAD samples and 41 normal colon tissues was performed by edgeR package (McCarthy et al., 2012; Robinson et al., 2010) and expressed as Fold-Change (FC) in the comparison tumour vs. normal. We denominated Variable Transcripts (VT) those transcripts showing statistically significant changes, at a False Discovery Rate (FDR) adjusted p-value < 0.05. We called positive transcripts (PositiveT) and negative transcripts (NegativeT) those VT having FC > 1 or FC < 1, respectively. The chromosomal distribution of the up-regulated transcripts (PositiveT) and downregulated transcripts (NegativeT) are reported in **Figure 58** (top panels), using the so-called Normalized Chromosomal Distribution Index (NCDI), as previously defined in Condorelli et al. (Condorelli et al., 2018).

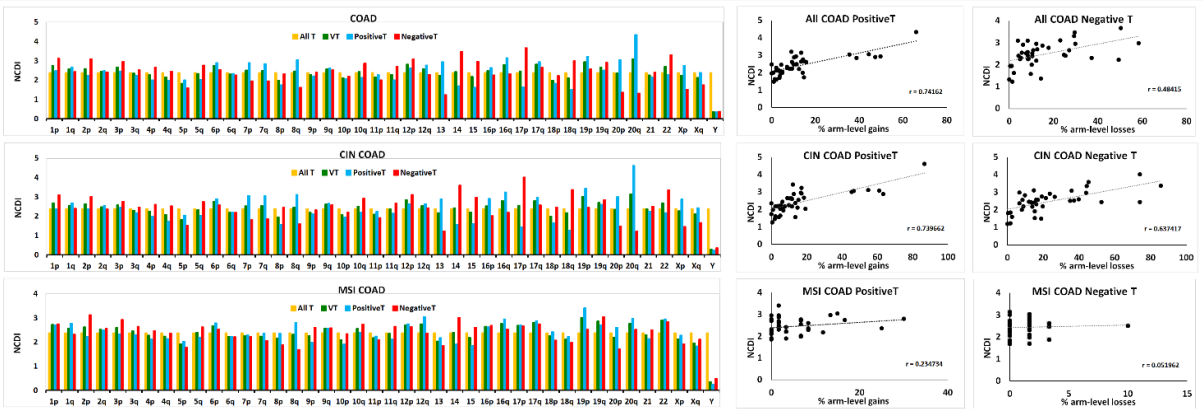


Figure 58. Percentage chromosomal distribution, expressed as NCDI (Normalized Chromosomal Distribution Index—see formula in Materials and Methods), of chromosomal density (number of specific transcripts divided by the total number of transcripts in that chromosome or chromosomal arm) of Variable Transcripts (VT), PositiveT and NegativeT. Values obtained for all transcripts (All T) located in a chromosome or chromosomal arm are also reported for comparison. Results obtained in COAD, CIN COAD and MSI COAD samples are shown in the left panels. Correlation plots of the NCDI values of PositiveT and NegativeT with the frequencies of arm-level gains and losses in each sample group are reported in the right panels. The Pearson's r index is reported in each correlation plot.

Interestingly, chromosomes that bear frequently arm-level gains (BCNGs), such as Chr20, 8q, 7, and 13 show high NCDI values of PositiveT, reflecting a relatively high chromosomal density of up-regulated transcripts (number of up-regulated transcripts divided by the total number of transcripts in that chromosomal region). On the other hand, chromosomes showing a high frequency of arm-level losses show high NCDI values of NegativeT, indicating a relatively high chromosomal density of downregulated transcripts. This is confirmed by a high Pearson's r correlation index between the NCDI values of PositiveT and the percentage of arm-level gains, or between NCDI values of NegativeT and the percentage of arm-level

losses (**Figure 58** top panels). However, it should be noted that 19p has a relatively high NCDI value of PositiveT, although bearing one of the lowest percentages of BCNGs (**Figure 57** and **Figure 58**).

Similar results have been obtained analysing 223 CIN COAD samples (**Figure 58** middle panels). On the basis of these results, it is not possible to establish if the high chromosomal density of up-regulated transcripts is a simple consequence of the high frequencies of BCNGs in the selected chromosomes, or also whether or not it is an intrinsic property of the selected chromosomes in cancer, even when they are not copy-number aberrant. On the contrary, no significant correlation of PositiveT or NegativeT with arm-level chromosomal aberrations is observed by analysing 60 MSI samples, in accordance with the fact that those tumours show an almost normal karyotype (**Figure 58** bottom panels).

4.2.3. Chromosomal Distribution of Arm-Level CNAs in "Selected COAD" Groups

We organized COAD samples in groups, called "Selected COAD" groups, bearing a specific BCNG. Each "Selected COAD group" was compared with a corresponding "Control COAD" group composed of tumours lacking any arm-level CNAs on the chosen chromosome. Among all 433 COAD samples we selected groups characterized by the following BCNGs: whole (w) Chr20-gain, i(20q), wChr8-gain, i(8q), wChr13-gain and wChr7-gain. Similar groups were also organized for CIN COAD samples (n = 223). In **Figure 59** the chromosomal distribution of arm-level gains and losses is shown for wChr8-gain and i(8q) in all tumour samples or in CIN samples. Since the "Chr8 Control group" (COAD samples without aberrations in Chr8, n = 153) contains many normal karyotype samples (such as MSI samples), the frequencies of chromosomal aberrations are very low (**Figure 59A**). On the other hand, the "Chr8 Control group" in CIN COAD samples (Chr8 control CIN group, n = 47, **Figure 59D**) shows higher frequencies of chromosomal aberrations with a profile similar to that observed in the wChr8-gain CIN group (except for Chr8 aberrations) (**Figure 59E**). In order to provide a quantitative score of the "profile similarity" between "control CIN groups" and "selected CIN groups", we performed a correlation analysis between the frequencies of arm-level gains and losses in each chromosomal arms of the two different groups (obviously excluding the value of the frequency of gains in the selected chromosome of each group, since these values are 0% in the control group and 100% in the gain group). We obtained the following Pearson's correlation coefficients: wChr20-gain vs. Chr20 control r = 0.83, i(20q) vs. Chr20 control r = 0.75,

wChr8-gain vs. Chr8 control r = 0.92, i(8q) vs. Chr8 control r = 0.93, wChr7-gain vs. Chr7 control r = 0.96, wChr13-gain vs. Ch13 control r = 0.96.

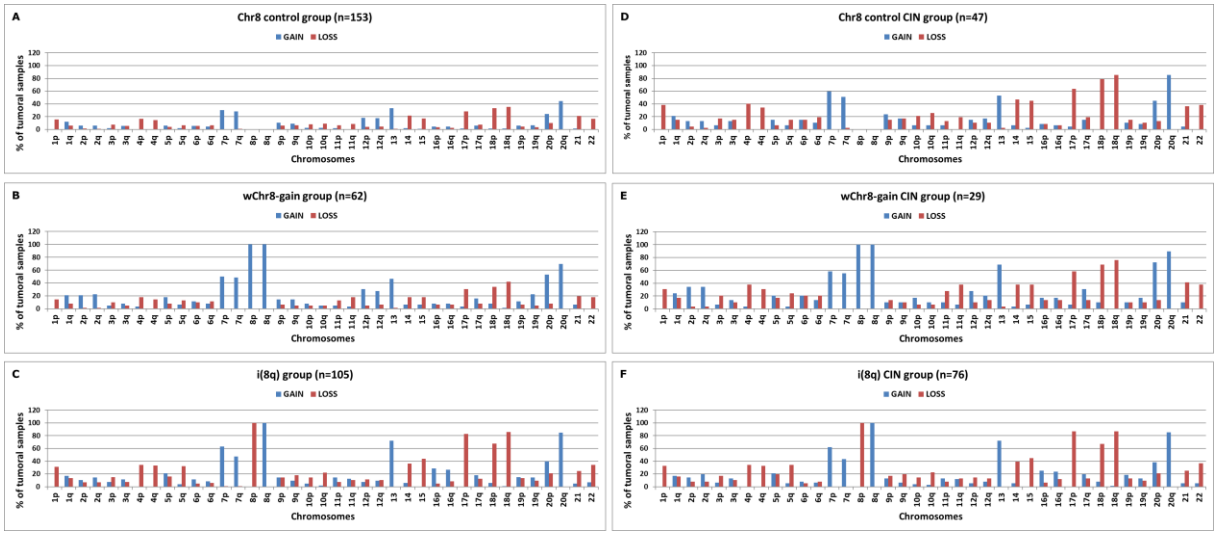


Figure 59. Arm-levels gains and losses in Chr8 control COAD group (A,D) and in wChr8-gain (B,E) and i(8q) (C,F) selected COAD groups. Results are expressed as a percentage of tumour samples bearing arm-level gains or losses in the p or q arm in all COAD samples (A–C) or in CIN COAD samples (D–F).

4.2.4. Transcriptome Analysis in "Selected COAD" Groups

For each COAD group, four different "indices" were used in order to estimate transcript level differences (Condorelli et al., 2018): (a) Linear fold-changes obtained comparing all COAD samples (Selected + Control COAD) vs. normal colonic mucosae (denominated FC1); (b) linear fold-changes obtained comparing COAD samples bearing a specific arm-level CNA (Selected COAD group) to samples not bearing it (Control COAD group) (denominated FC2); (c) linear fold-changes obtained comparing "Control COAD" group to normal colonic mucosae (denominated FC3); (d) linear fold-changes obtained comparing "Selected COAD" group to normal colonic mucosae (denominated FC4). The number of patients in each "Selected COAD" and corresponding "Control COAD" group are reported in **Figure 59**.

We called OverT the VT expressing FC2 > 1.3, representing the transcripts that are overexpressed in the "selected COAD group" in comparison to the corresponding "control COAD group". Finally, we called PositiveT those VT having FC3 > 1 and FC4 > 1, and NegativeT those VT expressing FC3 < 1 and FC4 < 1. In other words, PositiveT and NegativeT are transcripts that are up-regulated or downregulated, respectively, in both the "selected COAD group" and the "control COAD group", in comparison to normal colon tissue.

However, in the results shown in **Figure 60A**, PositiveT have been further subdivided in two categories: PositiveT-FC3 and PositiveT-FC4. PositiveT-FC3 represent transcripts up-

regulated in the "control COAD group" in comparison to normal colon tissue (T having FC3 > 1 and FDR < 0.05 in the comparison "Control COADs" vs. normal colonic mucosae), while PositiveT-FC4 are transcripts up-regulated in "selected COAD group" (T having FC4 > 1 and FDR < 0.05 in the comparison "Selected COADs" vs. normal colonic mucosae). We calculated the percentage distribution of PositiveT-FC3 and PositiveT-FC4, expressed as NCDI values, for each chromosome arm or acrocentric chromosome.

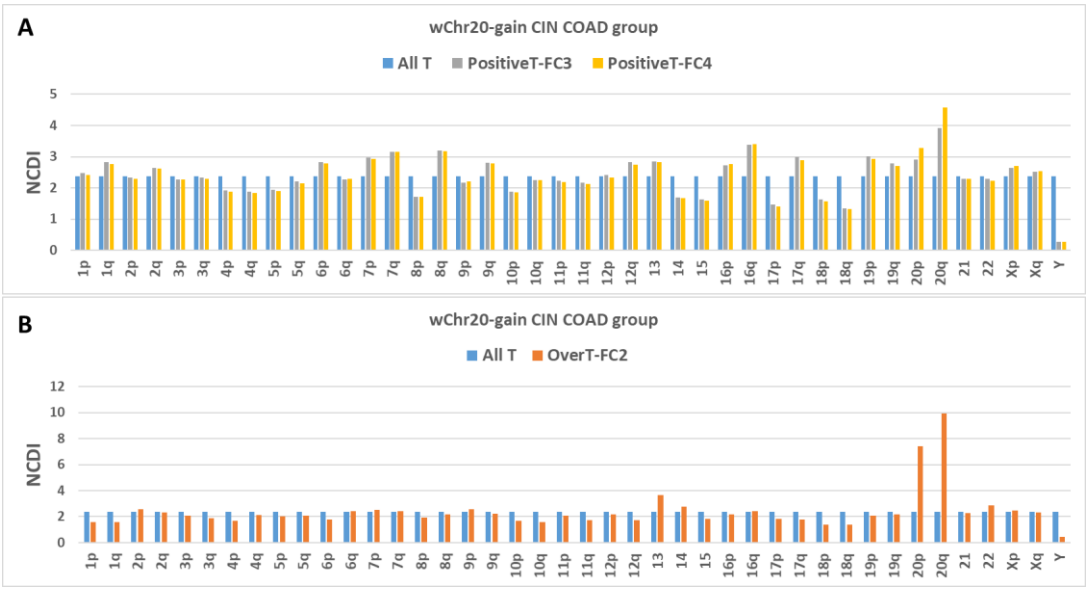


Figure 60. Percentage chromosomal distribution (expressed as NCDI) of different transcript classes in Chr20 control and wChr20-gain CIN-positive colon cancer groups: (A) All transcripts (All T), PositiveT-FC3, and PositiveT-FC4); (B) All T and OverT-FC2.

Interestingly, the NCDI of PositiveT-FC4 is significantly higher than corresponding values of PositiveT-FC3 in all analysed selected chromosomes (i.e., Chr20 in 60A and Chr20q, Chr8, Chr8q, 13 and 7 in other selected COAD groups as reported in Supplementary Figure S2). Moreover, the NCDI values of PositiveT-FC3 in selected chromosomes are not significantly different from the NCDI of the entire repertoire of transcripts, with the exception of a slight increase in Chr20q (**Figure 60A**). These data suggest that the relatively high density of up-regulated genes in selected chromosomes (Chr20, 8, 13 and 7) is mainly the consequence of the presence of BCNGs. Moreover, it should be noted that some chromosomes, such as Chr19 and Chr16, have a relatively high NCDI value, but do not show a high frequency of BCNGs. Therefore, a high density of cancer upregulated genes in disomic chromosomes cannot be considered one of the major predisposing factors for the development of the gain aberrations of those chromosomes during cancer progression.

The number of transcripts bearing FC values higher than a predetermined threshold (FC2 > 1.3, FC3 > 1, FC4 > 1) are reported in the Venn diagrams in **Figure 61** (wChr8-gain group). The transcripts contained in the region of overlapping of the three circles of the Venn diagram

(corresponding to transcripts with FC2 > 1.3, FC3 > 1 and FC4 > 1) are called "Over-PositiveT", because these transcripts are overexpressed when comparing the "wChr8-gain group", with its corresponding "Chr8 Control group" (FC2 > 1.3), and show positive values of FC3 and FC4 (**Figure 61**). A high percentage of Over-PositiveT, as shown by the NCDI value, is localized in the gained chromosome characterizing the selected group (bottom of **Figure 61**, showing NCDI values for each of the chromosomal arms and acrocentric chromosome). Moreover, the NCDI value is lower for Over-NegativeT, defined as overexpressed transcripts with FC2 > 1.3, but showing FC3 < 1 and FC4 < 1 (**Figure 61**, and **Figure 62** for CIN COAD). The results of the hypergeometric test confirm an enrichment of OverT among PositiveT and/or their depletion among NegativeT (**Table 21**).

	Chr20p	Chr20q	Chr8p	Chr8q	Chr13	Chr7p	Chr7q
Fold-enrichment of OverT among PositiveT (Hypergeometric *p*-value)	1.11 (0.00040)	1.00 (0.49343)	1.16 (0.00109)	1.08 (0.02179)	1.12 (2.93×10^{-7})	1.10 (0.02120)	1.05 (0.14272)
Fold-depletion of OverT among NegativeT (Hypergeometric *p*-value)	−2.69 (4.55×10^{-11})	−1.6 (3.30×10^{-7})	−1.45 (0.00109)	−1.19 (0.01940)	−1.8 (5.94×10^{-11})	−1.38 (0.00331)	−1.3 (0.00653)
Fold-enrichment of OverT among Gained VEL-T (Hypergeometric *p*-value)	2.14 (2.23×10^{-6})	1.45 (0.00021)	1.99 (0.03708)	3.29 (1.87×10^{-13})	2.25 (2.14×10^{-7})	2.33 (1.85×10^{-6})	1.76 (0.00151)
Fold-enrichment of OverT among Fitness-T (Hypergeometric *p*-value)	2.43 (5.03×10^{-7})	2.22 (5.55×10^{-15})	4.41 (2.31×10^{-13})	3.14 (1.5×10^{-12})	3.58 (5.81×10^{-26})	2.61 (4.88×10^{-6})	1.65 (0.00929)
Fold-enrichment of OverT among AmpT (Hypergeometric *p*-value)	9.75 (6.56×10^{-6})	1.15 (0.109)	3.73 (0.250)	NC [1]	2.32 (0.002)	NC [1]	1.07 (0.523)

[1] not computable for the absence of AmpT.

Table 21. Fold-enrichment of OverT among different classes of transcripts evaluated by hypergeometric test

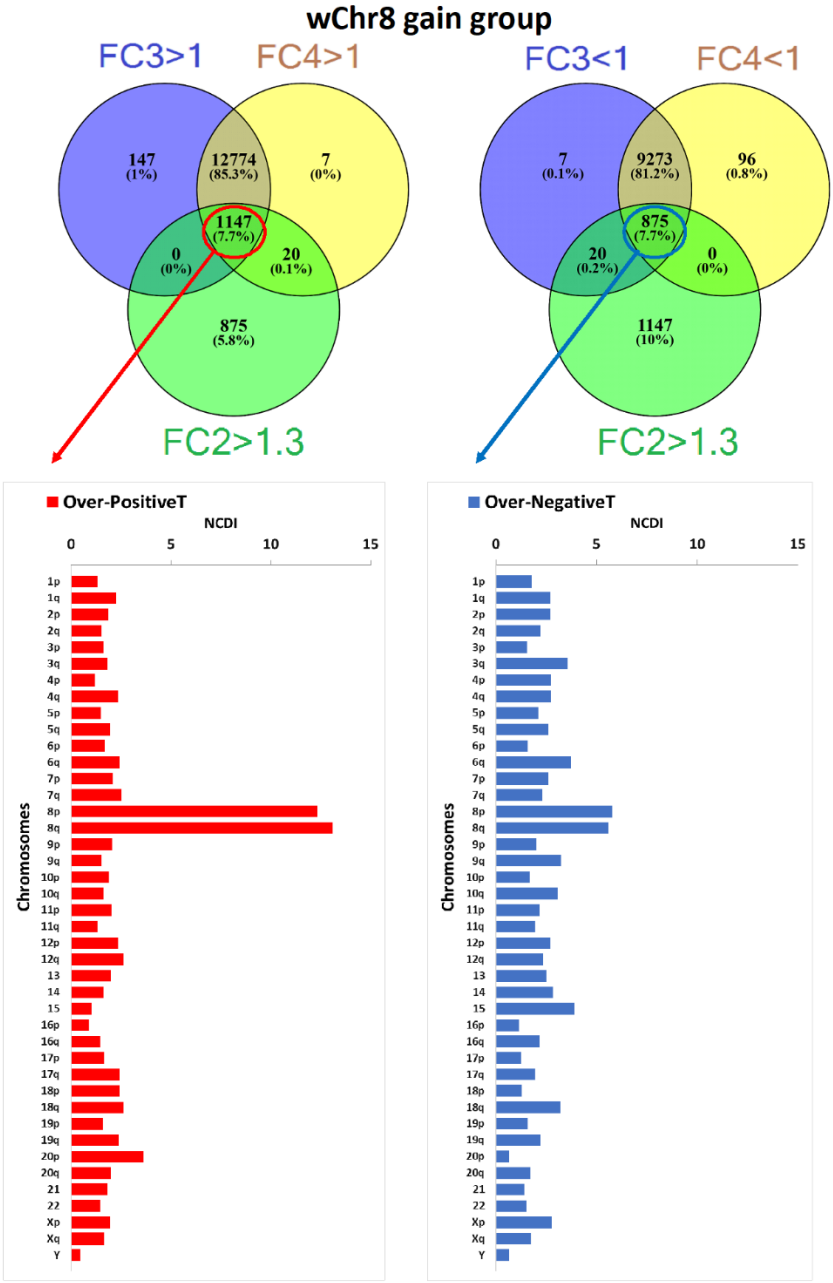


Figure 61. (Top) Venn Diagrams showing the number of transcripts expressing levels higher or lower than an established fold-change value. (Bottom) Normalized Chromosomal Distribution Index (NCDI) of Over-PositiveT (transcripts with FC2 > 1.3, FC3 > 1, FC4 > 1) or Over-NegativeT (transcripts with FC2 > 1.3, FC3 < 1, FC4 < 1) in all chromosomal arms and acrocentric chromosomes. Results have been obtained by analysing the Chr8 control COAD group, wChr8-gain COAD group and normal colonic mucosae.

These data suggest a differential susceptibility of PositiveT and NegativeT to gene dosage effects. Bearing in mind that positive transcripts are up-regulated transcripts in comparison to normal tissue, and that negative ones are down-regulated transcripts, this phenomenon has been denominated the "positive caricature effect", based on the definition of "caricature" as an exaggeration of tumour's up-regulated gene expression features (Condorelli et al., 2018). Of course, this is just a statistical observation, suggesting that the probability of gene-dosage

effects in down-regulated genes is lower than in up-regulated ones. Each selected COAD group's three main transcript types were evaluated according to the annotation of GRCh38.p12 provided by the Ensembl BioMart tool (www.ensembl.org/biomart/): "Protein-coding", "Non-Coding" and "lincRNA" (Long intergenic non-coding RNA). In the non-coding group, the TEC (To be Experimentally Confirmed) and NA (Not Assessed) entries were excluded. As shown in **Figure 62**, all three transcript types show an increase of the NCDI values of OverT and Over-PositiveT in the chromosome characterizing the "selected COAD group", although the effect is more evident when analysing protein-coding transcripts. Moreover, Over-PositiveT NCDI values are significantly higher than Over-NegativeT ones in all transcript types.

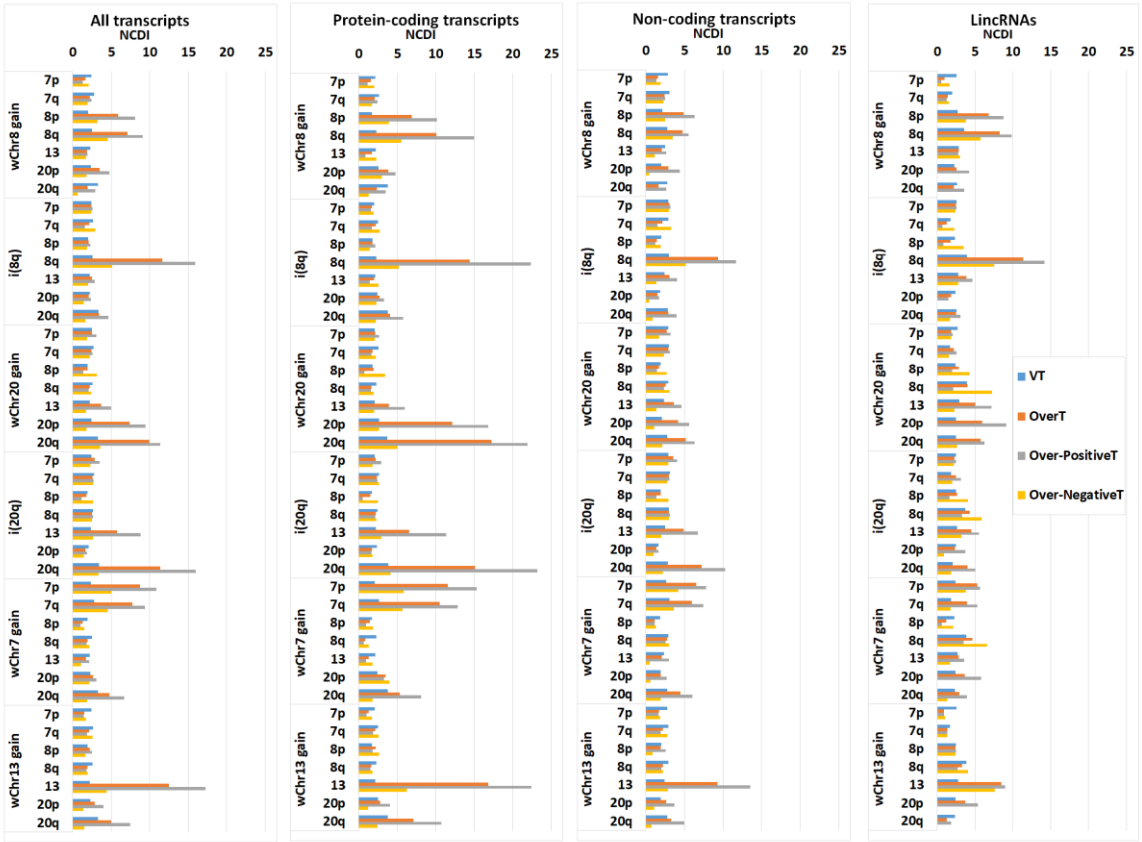


Figure 62. NCDI of different transcript classes (VT, OverT, Over-PositiveT, Over-NegativeT) in selected chromosomal arms or acrocentric chromosomes in different groups of CIN COAD samples. Calculation of NCDI values are reported for all transcript types or, separately, for "Protein-coding", "Non-Coding" and "lincRNA".

In order to show that the difference between Over-PositiveT and Over-NegativeT is not simply dependent on the transcript level, transcripts were subdivided into different sets or bins according to their TPM (Transcripts Per Million) values, and NCDI values in Chr8q were calculated for each bin in the wChr8-gain COAD group (**Figure 63**). Although an increase of NCDI can be observed at higher TPM values, the NCDI value of Over-NegativeT is lower

than that of overexpressed transcripts (OverT; FC2 > 1.3) or Over-PositiveT at all examined TPM bins.

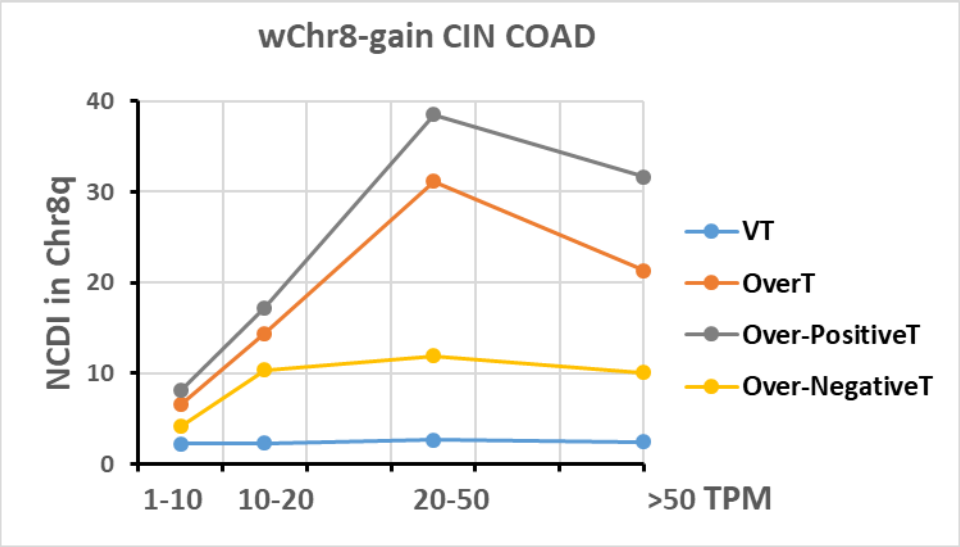


Figure 63. Transcripts are subdivided into different bins according to TPM values as indicated in the x-axis. The NCDI value in Chr8q is reported in the y-axis. Data obtained in the wChr8-gain CIN COAD group are reported.

4.2.5. Recurrent Gained Variant Enhancer Loci

Colon carcinogenesis is accompanied by locus-specific gains and losses of enhancer activity, called "Variant Enhancer Loci" (VELs) (Akhtar-Zaidi et al., 2012; Cohen et al., 2017). Cohen et al. have performed high-resolution H3K27ac ChIP-seq profiles in seven specimens of normal colonic epithelial crypts and 35 colorectal cancer samples(Cohen et al., 2017). DNA sites in which the H3K27ac mark was more enriched in colorectal samples than in the normal crypts were termed "gained VELs". Moreover, Cohen et al. have assigned VELs to their putative target genes and corresponding transcripts, by an experimentally validated computational method that predicts enhancer-gene interactions(Cohen et al., 2017). Gained VELs that are present in 10 or more colorectal cancer samples were deemed as significantly recurrent, and transcripts linked to recurrent gained VELs (here denominated "gained VEL-T") are up-regulated in primary colorectal cancer compared to normal tissue (Cohen et al., 2017). We report here that OverT are significantly enriched among gained VEL-T transcribed in selected chromosomal arms of selected COAD CIN groups (**Table 25**).

As shown in **Figure 64**, the NCDI value of Gained VEL-OverT is more elevated than that of OverT in chromosome arms 20p, 20q, 8q, 13, 7p and 7q, suggesting that epi-genomic changes associated with gained VEL may play a role in the susceptibility to the transcriptional dosage cis-effect. In several chromosomes (20p, 20q, 8q, 7p) the NCDI increases are higher than those observed in Over-PositiveT.

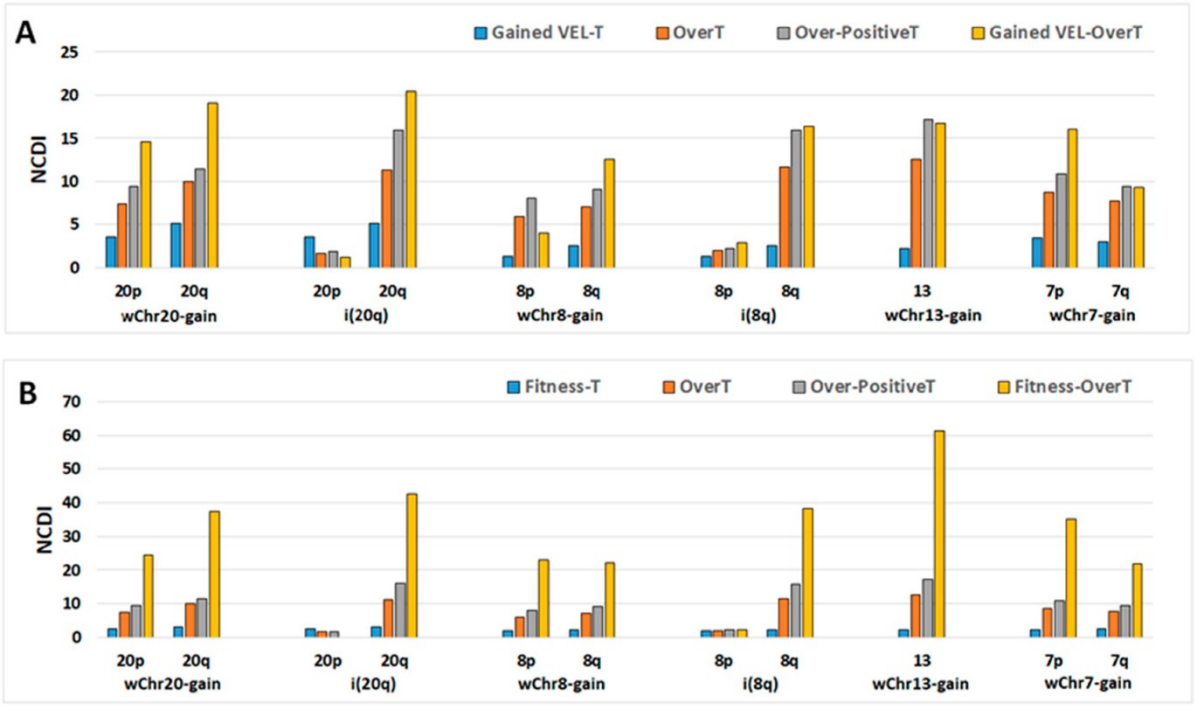


Figure 64. NCDI values of Gained Variant Enhancer Loci-Transcripts (VEL-T), OverT, Over-PositiveT, and Gained VEL-OverT (A), and Fitness-T, OverT, Over-PositiveT and Fitness-OverT (B) have been calculated in selected chromosomes of each Selected COAD group as indicated in the x-axis.

4.2.6. Cancer Fitness Genes

In order to identify genes that are required for cancer cell fitness, genome-scale CRISPR-Cas9 screens have been performed in cancer cell lines (Behan et al., 2019). In particular, Behan et al. have identified genes required for cell growth or viability in colon cancer cell lines, whose transcripts are here abbreviated as "Fitness-T" (Behan et al., 2019).

As shown in **Table 25**, OverT are significantly enriched among Fitness-T transcribed in selected gained chromosomal arms of selected COAD CIN groups. Moreover, NCDI values of Fitness-OverT are significantly higher than those of OverT and Over-PositiveT in selected chromosomal arms of each COAD group (**Figure 64**).

The Venn diagrams show the shared genes among Over-positiveT, Fitness-OverT and Gained VEL-T gene lists.

4.2.7. Recurrent Focal Amplifications

Sack et al. have reported a list of genes included in recurrent focal amplification in colorectal adenocarcinomas (Sack et al., 2018). Although the focus of the present work is on arm-level gains or BCNGs, we wondered whether transcripts associated with focal amplifications (AmpT) are enriched among OverT genes identified in BCNG regions. As shown in **Table 25**, a trend towards an enrichment was observed in some chromosomes (20p, 8p, 13), but these results, with the exception of chr13, are not significant. Interestingly, some chromosome arms frequently involved in BCNGs do not show recurrent focal amplifications (8q, 7p), thus suggesting that fundamental differences distinguish the cancer driver mechanisms of broad and focal chromosomal aberrations.

4.2.8. Ingenuity Pathway Analysis: Over-Positive T and Fitness-OverT

Ingenuity Pathway Analysis (IPA®) was performed using, as datasets, the lists of Over-PositiveT or Fitness-OverT obtained for each selected COAD group and the FC2 values. Each list of Over-PositiveT or Fitness-OverT contained only genes localized on a selected chromosome or chromosomal arm. In order to exploit data deriving from wChr-gain groups and isochromosomes, only genes located on the q arm of Chr20 and Chr8 were analysed. The IPA Core analysis identified several significant Canonical Pathways linked to cancer processes, such as pathways involved in proliferation and cell cycle control, cancer signalling, DNA repair, and amino acid metabolism. Proteins involved in the control of the cell cycle, such as E2F1 and RBL1 in Chr20, and CCNE2 and E2F5 in Chr8, are among the main determinants of such results.

In order to explore the functional interactions between genes located in different chromosomes, we also prepared a combined list of Over-PositiveT or FitnessT-OverT genes located on Chr20q, 8q, 13 and 7. Results of IPA core analysis (**Figure 65**) revealed 39 and 44 significant canonical pathways for Over-positiveT and Fitness-OverT, respectively. Twenty-two canonical pathways are shared between the two transcript classes (in red letters in **Figure 65).**

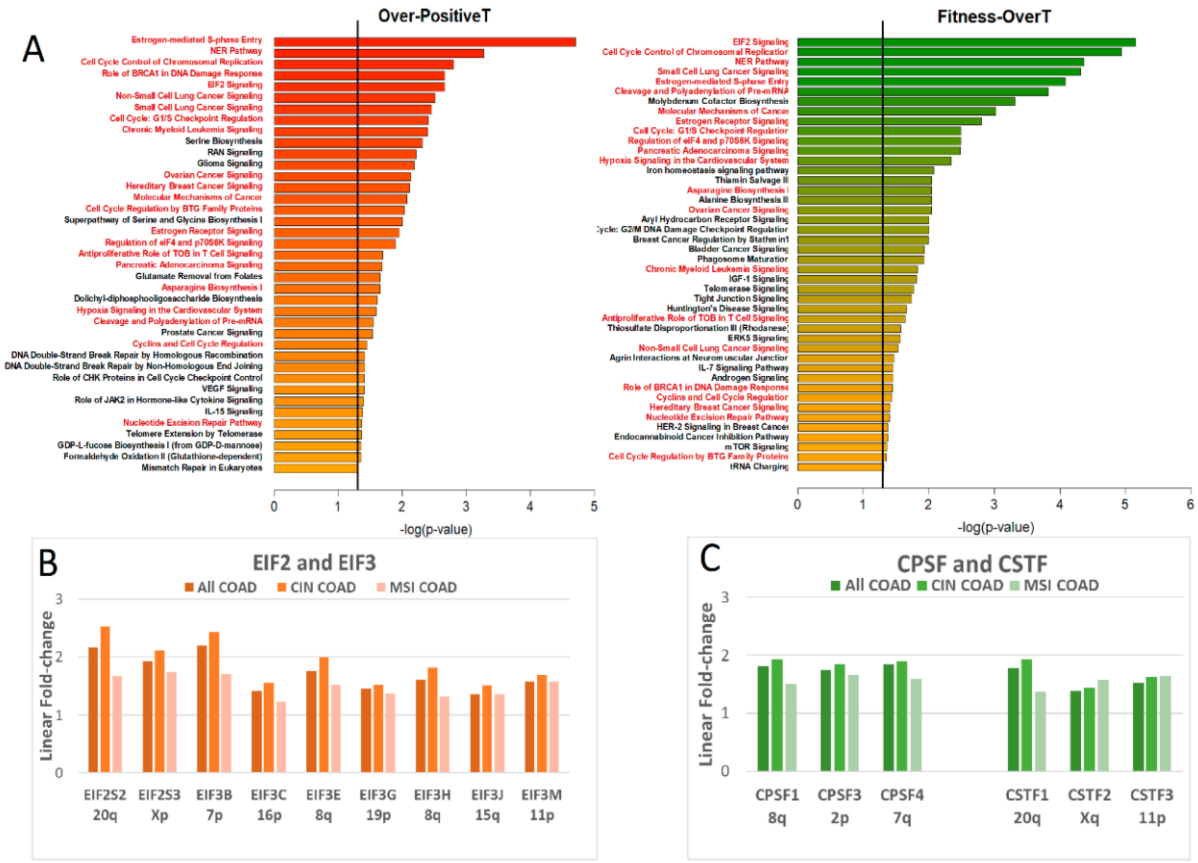


Figure 65. (A) The p-values of Canonical pathways significantly associated with Over-PositiveT (left) or Fitness- OverT (right) genes located in 20q, 8q, 13, 7. The vertical black line indicates a p-value equal to 0.05. Canonical pathways shared between Over-PositiveT and Fitness-OverT are written in red letters. (B) Expression level of some significantly up-regulated EIF2 and EIF3 genes in COAD samples. (C) Expression level of some significantly up-regulated CPSF and CSTF genes in COAD samples. In B and C, the expression level is expressed as linear fold-change in the comparison tumour vs. normal. Only genes significantly up-regulated (linear fold change >1.5 and False Discovery Rate (FDR) p-value < 0.05 in all COAD samples) are reported. Average values of 433 COAD samples, 223 CIN Coad and 60 MSI COAD are shown.

EIF2 signalling is one of the most significant pathways (top position for Fitness-OverT and 4th position for Over-PositiveT), with an activation Z-score > 2, indicating that the expression pattern of our dataset is consistent with the canonical pathway having more activity, according to the Ingenuity Knowledge Base®.

EIF2S2 (chr20q11.22), EIF3B (chr7p22.3), EIF3E (Chr8q23.1) and EIF3H (chr8 q24.11) are among the genes in our dataset that contribute to the identification of this EIF2 signalling pathway. In **Figure 65B** we report the expression levels of different EIF2 and EIF3 genes showing a significant increase in tumour versus normal tissue (linear fold change >1.5, FDR p-value < 0.05) in all COAD or in CIN and MSI COAD samples. Interestingly, EIF2S2, EIF2S3, EIF3B and EIF3E show the largest increase in expression in comparison to normal

tissue. Moreover, the expression of several EIF2/3 genes in CIN tumours is higher than that of MSI ones.

“Cleavage and Polyadenylation of Pre-mRNA” is another significant pathway, in both Over-PositiveT and Fitness-OverT, which contains subunits of multi-protein complexes. One of these multi-subunit complexes is the cleavage and polyadenylation specificity factor (CPSF) playing a role in the 3′ processing of pre-mRNAs by the recognition of the AAUAAA signal and by the interaction with other complexes and enzymes involved in both RNA cleavage and poly(A) synthesis. CPSF includes the proteins CPSF1 (also known as CPSF160), CPSF2 (CPSF100), CPSF3 (CPSF73), CPSF4 (CPSF30), FIP1L1 and WDR33. Another multi-subunit complex belonging to the “Cleavage and Polyadenylation of Pre-mRNA pathway” is the “cleavage stimulation factor (CSTF)”, a trimer of CSTF1 (CstF50), CSTF2 (Cstf64) and CSTF3 (CstF77)(Mandel et al., 2008; Shi et al., 2009; Tian & Manley, 2016). In **Figure 65C** we report the expression levels of different genes of the CPSF and CSTF complexes, showing a significant increase in tumour versus normal tissue (linear fold change > 1.5, FDR p-value < 0.05) in all COAD or in CIN and MSI COAD samples. A significant difference between CIN and MSI COADs is observed for genes located in chromosomes undergoing frequent BCNGs, such as CPSF1 (chr8q24.3), CPSF4 (chr7q22.1) and CSTF1 (Chr20q13.2). Indeed, CPSF1, CPSF4 and CSTF1 are among genes in our dataset of Over-PositiveT and Fitness-OverT that contribute to the identification of the pathway “Cleavage and Polyadenylation of Pre-mRNA”.

4.3 GLIOBLASTOMA MULTIFORME (GBM)

This work of thesis is also focused on the preliminary investigation of chromosome 7-gain in cancers as well as on understanding of EGFR role with (7p-FA+) or without focal amplification (7p-FA-) in those GBM tumours bearing the chromosome 7-gain. Usually, chr7-gain and the 7p-FA+ can occur separately or jointly and it is not yet clear if their main pathogenic mechanisms are largely overlapping or if major mechanistic differences characterize the two aberrations. In order to shed light on these issues, in the present work we compared cytogenomic, transcriptomic and clinic-pathological data of GBMs bearing the 7p-FA and/or chr7-gain aberrations or not-bearing both aberrations.

4.3.1 Chromosome 7-gain analysis in 33 cancer types from TCGA studies

To investigate the status of chromosome 7 in several cancer types, we download the cytogenetics data from TCGA concerning the whole project based on 33 cancer types. The details of the cancer types and number of samples are showed in **Table 22.** To give a dimension of analysis 11295 tumours samples are analysed. For each cancer types the followed features are calculated 1) number of samples bearing 7p-gain, 2) percentage of samples bearing 7p-gain 3) number of samples bearing 7q-gain 4) percentage of samples bearing 7q -gain 5) number of samples with whole 7-gain 6) percentage of samples bearing 7-gain (% trisomy of wchr7) 7) number of samples with disomic chromosome 7 8) percentage of samples with disomic chromosome 7 and 9) Total number of SNPs array data.

The result about the percentage of samples bearing gain of chromosome 7 (% trisomy of wchr7) in 33 TCGA cancer identified GBM as the first tumours characterized by the highest percentage of 7-gain (67% trisomy of wchr7), followed by KIRP (55.47%), COAD CIN (46.90%), READ CIN (45.09%).

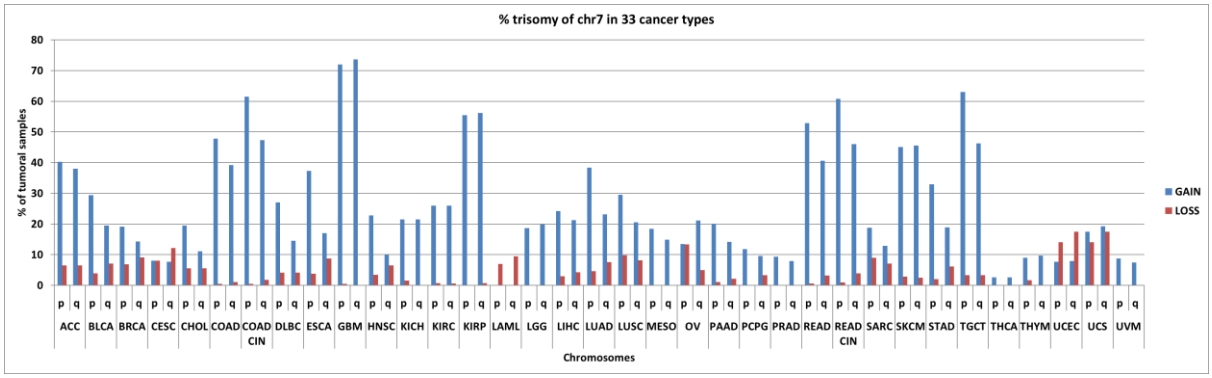


Figure 66: Percentages of chromosome 7 (gain and loss) in 33 cancer types.

The **figure 66** shows the relation of percentage of p and q chromosomal arms and the cancer types by considering the gain and the loss of chromosome 7. GBM shows the highest percentage of gain for individual chromosomal arms (7p and 7q). Uterine corpus endometrial carcinoma (UCEC), Uterine carcinosarcoma (UCS) and Acute myeloid leukaemia (LAML) show an interesting percentage of loss for individual chromosomal arm (7p and 7q).

	n.samples with tris 7p	p arm % Gain	n.samples with tris 7q	q arm % Gain	n.sample with tris wchr7	% trisomy of wchr7	n.samples with disomy of wchr7	% disomy of wchr7	SNPs array tot data
GBM	426	71.9594595	436	73.64864865	401	67.73648649	79	13.34459459	592
KIRP	157	55.4770318	159	56.18374558	157	55.4770318	112	39.57597173	283
COAD CIN	139	61.5044248	107	47.34513274	106	46.90265487	73	32.30088496	226
READ CIN	62	60.7843137	47	46.07843137	46	45.09803922	35	34.31372549	102
TGCT	94	63.0872483	69	46.30872483	65	43.62416107	33	22.14765101	149
READ	82	52.9032258	63	40.64516129	61	39.35483871	66	42.58064516	155
COAD	210	47.8359909	172	39.17995444	170	38.72437358	198	45.10250569	439
SKCM	202	45.0892857	204	45.53571429	172	38.39285714	139	31.02678571	448
ACC	37	40.2173913	35	38.04347826	33	35.86956522	39	42.39130435	92
KIRC	133	25.9765625	133	25.9765625	126	24.609375	328	64.0625	512
KICH	14	21.5384615	14	21.53846154	14	21.53846154	46	70.76923077	65
LUAD	217	38.3392226	131	23.14487633	111	19.61130742	178	31.44876325	566
LGG	96	18.6770428	102	19.84435798	91	17.70428016	342	66.53696498	514
LIHC	90	24.1935484	79	21.23655914	65	17.47311828	200	53.76344086	372
STAD	145	32.9545455	83	18.86363636	74	16.81818182	186	42.27272727	440
MESO	16	18.3908046	13	14.94252874	13	14.94252874	58	66.66666667	87
BLCA	121	29.4403893	80	19.46472019	59	14.35523114	183	44.52554745	411
LUSC	144	29.5687885	100	20.5338809	69	14.16837782	161	33.05954825	487
ESCA	68	37.3626374	31	17.03296703	25	13.73626374	44	24.17582418	182
PAAD	37	20.1086957	26	14.13043478	25	13.58695652	115	62.5	184
DLBC	13	27.0833333	7	14.58333333	6	12.5	28	58.33333333	48
BRCA	208	19.1881919	155	14.29889299	125	11.53136531	549	50.64575646	1084
CHOL	7	19.4444444	4	11.11111111	4	11.11111111	25	69.44444444	36
OV	79	13.5042735	124	21.1965812	53	9.05982906	115	19.65811966	585
PCPG	21	11.7977528	17	9.550561798	16	8.988764045	130	73.03370787	178
THYM	11	8.9430894	12	9.756097561	11	8.943089431	92	74.79674797	123
SARC	48	18.8235294	33	12.94117647	20	7.843137255	116	45.49019608	255
HNSC	119	22.7533461	52	9.942638623	41	7.839388145	257	49.13957935	523
UVM	7	8.7500000	6	7.5	6	7.5	73	91.25	80
PRAD	46	9.3117409	39	7.894736842	35	7.085020243	394	79.75708502	494
UCS	10	17.5438596	11	19.29824561	4	7.01754386	21	36.84210526	57
UCEC	41	7.7504726	42	7.939508507	29	5.482041588	377	71.26654064	529
CESC	24	8.0808081	23	7.744107744	10	3.367003367	192	64.64646465	297
THCA	13	2.6000000	13	2.6	13	2.6	449	89.8	500
LAML	0	0.0000000	0	0	0	0	99	49.5	200

Table 22: Percentage of trisomy 7 in 33 cancer types

In order to investigate the transcriptomic state of chromosome 7 gain in GBM in comparison to the other cancer types, a hierarchical clustering was performed. In **figure 67** the logarithm of Fold change (defined FC2) was evaluated. As mentioned previously, FC2 concerns comparison between samples bearing a specific arm-level aberration (Selected GBM group) to samples not bearing it (Control GBM group) in accordance with the model in **figure 4**. These bioinformatics pipelines were performed individually for each cancer types. Data are clustered by the unweighted pair group method with arithmetic mean (UPGMA) with Euclidean distance. In the analysis are included only those transcripts with an high expression ($FC2 > 1.3$) and high significance (adj p value < 0.05) in GBM cancer.

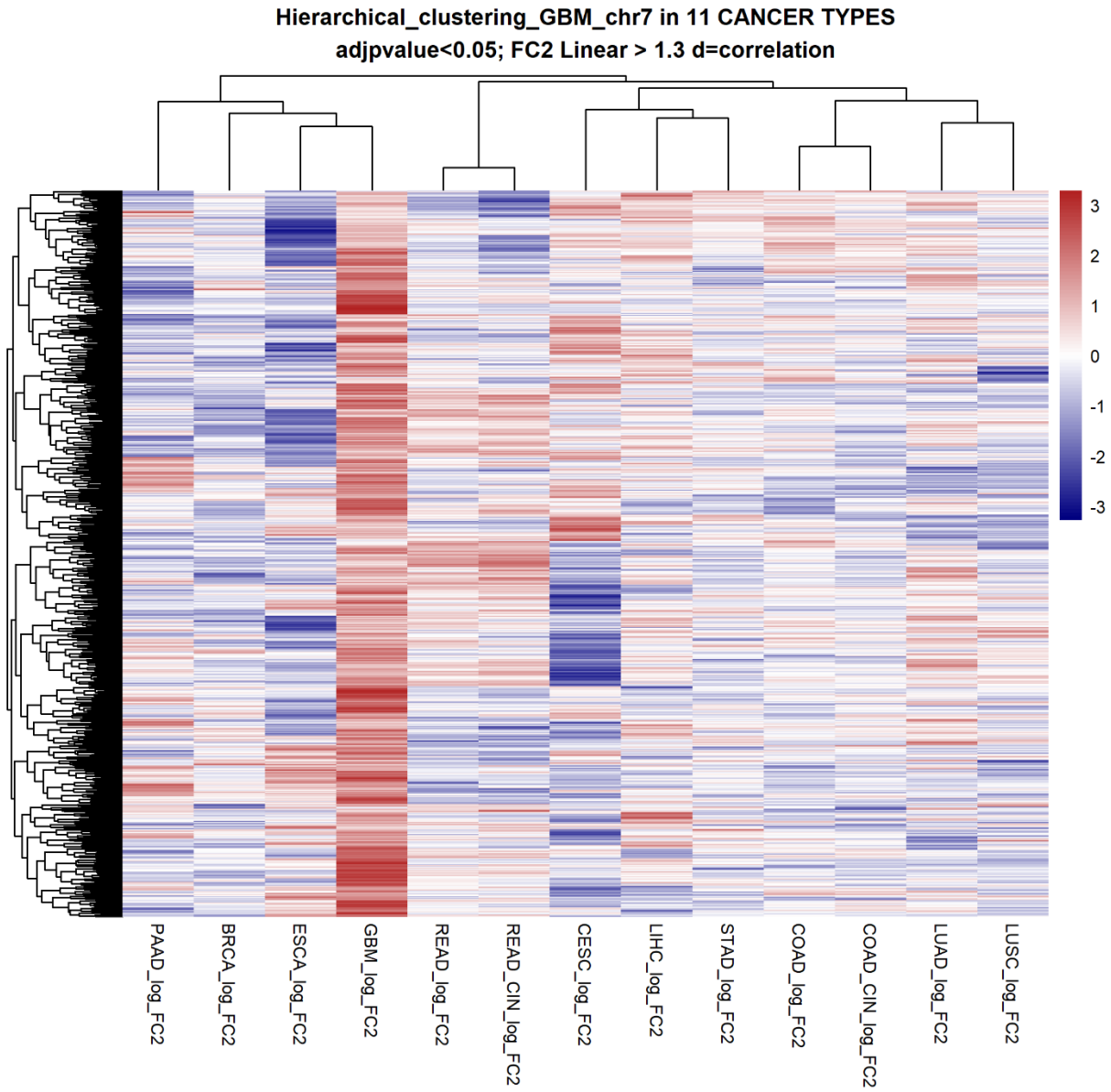


Figure 67: Unsupervised hierarchical clustering of 13 cancer types.

In GBM, the over-expression of those genes located on chromosome 7 is a specific expression pattern not comparable with other cancer types. In fact, they show a clear cut-off in terms of expression pattern for those gene located on chromosome 7.

These preliminary results lead to a deep analysis of pathogenetic role of chromosome 7 in GBM.

4.3.2 Chromosome 7-gain in GBM: Cytogenetics analysis

To investigate the role of chromosome 7 gain in the tumorigenesis of glioblastomas, we performed a comparative and integrative bioinformatics analysis involving cytogenetic SNP array data (592 samples), RNA-seq data (174 samples in total subdivided in 128 primary tumour and 5 normal tissue) and whole exome sequencing (WES) data (345 samples) from The Cancer Genome Atlas (TGCA).

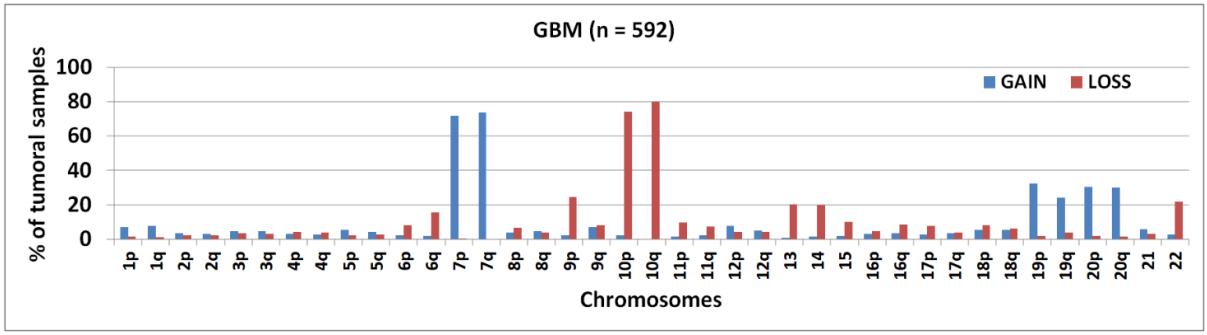


Figure 68: Percentage of samples bearing chromosomal arm-level gains or losses in GBM samples

The **figure. 68** shows the frequencies of chromosomal arm-level aberrations for the entire series of 592 GBM samples from TCGA. The gain of chromosome (chr) 7 (p: 72%; q: 74%) and the loss of chr10 (p:74%, q:80%) are more common abnormalities. Other aberrations, with a frequency higher than 15%, are gains of chr19 (p: 32%; q: 24%) and chr20 (p: 30%; q: 30%), and losses of chr9p (25%), chr6q (16%) and acrocentric chromosomes 13 (20%), 14 (20%), 22 (22%).

In order to confirm the association of chr19-gain with 7pFA in a larger number of samples, we exploited the EGFR gene copy number analysis (GISTIC data on EGFR gene copy number retrieved from cBioPortal, www.cbioportal.org/, on April 13 2021) in the entire cohort of 522 samples profiled by SNP array (Brennan et al., 2013). As shown in **Table 29**, such analysis confirmed the significant association between chr19-gain and 7pFA. A similar association was found for chr20-gain, chr10-loss and chr7-gain. although the strength of association is higher for chr19-gain (reported as relative risk in **Table 29**).

	Putative EGFR region Copy Number Variation (n = 522)			
Only samples with SNP array data (n =592)	**7p-FA+**	**7pFA-**	**X^2 P-value**	**Relative Risk (95% CI)**
Chr7-gain (n = 353)	159	194	0.0053	1.63 (1.14 - 2.42)
Chr7-disomy (76)	21	55		
Chr10-loss (n = 387)	203	184	<0.0001	∞
Chr10-disomy (n = 50)	0	50		
Chr19-gain (n = 120)	86	34	<0.0001	2.071 (1.69 - 2.53)
Chr19-disomy (n =263)	91	172		
Chr20-gain (n = 146)	92	54	<0.0001	1.71 (1.41 - 2.06)
Chr20-disomy (n = 321)	118	203		
	Chr7gain / 7pFA-	**Chr7disomy / 7pFA-**		
Chr10-loss (n =184)	152	17	<0.0001	1.71 (1.34 - 2.39)
Chr10-disomy (n = 50)	22	20		
Chr19-gain (n = 34)	32	0	0.001	1.357(1.20 - 1.58)
Chr19-disomy (n =172)	115	41		
Chr20-gain (n = 54)	49	2	0.0008	1.29 (1.14-1.43)
Chr20-disomy (n = 203)	131	45		

Table 29 Chi-square association test for different chromosomal aberrations and 7pFA. SNP array total data

In order to study the transcriptional effects of aberrations of chr7, alone or in combination, we retrieved only samples profiled for gene expression (RNA-Seq) and copy-number status (SNP-array) and defined different subgroups of GBMs according to the copy number status of that chromosome. Moreover, the presence or absence of focal amplification of EGFR (as reported in Brennan et al 2013; supplementary table 5) was used as an additional parameter to differentiate GBM samples in 7pFA+ (EGFR focal amplification positive) and 7pFA- (EGFR focal amplification negative). According to those data we could distinguish 4 groups characterized by the presence of chr7-gain or chr7-disomy and the positivity or negativity for EGFR focal amplification (7pFA+ or 7pFA-) (**Figure 69**)

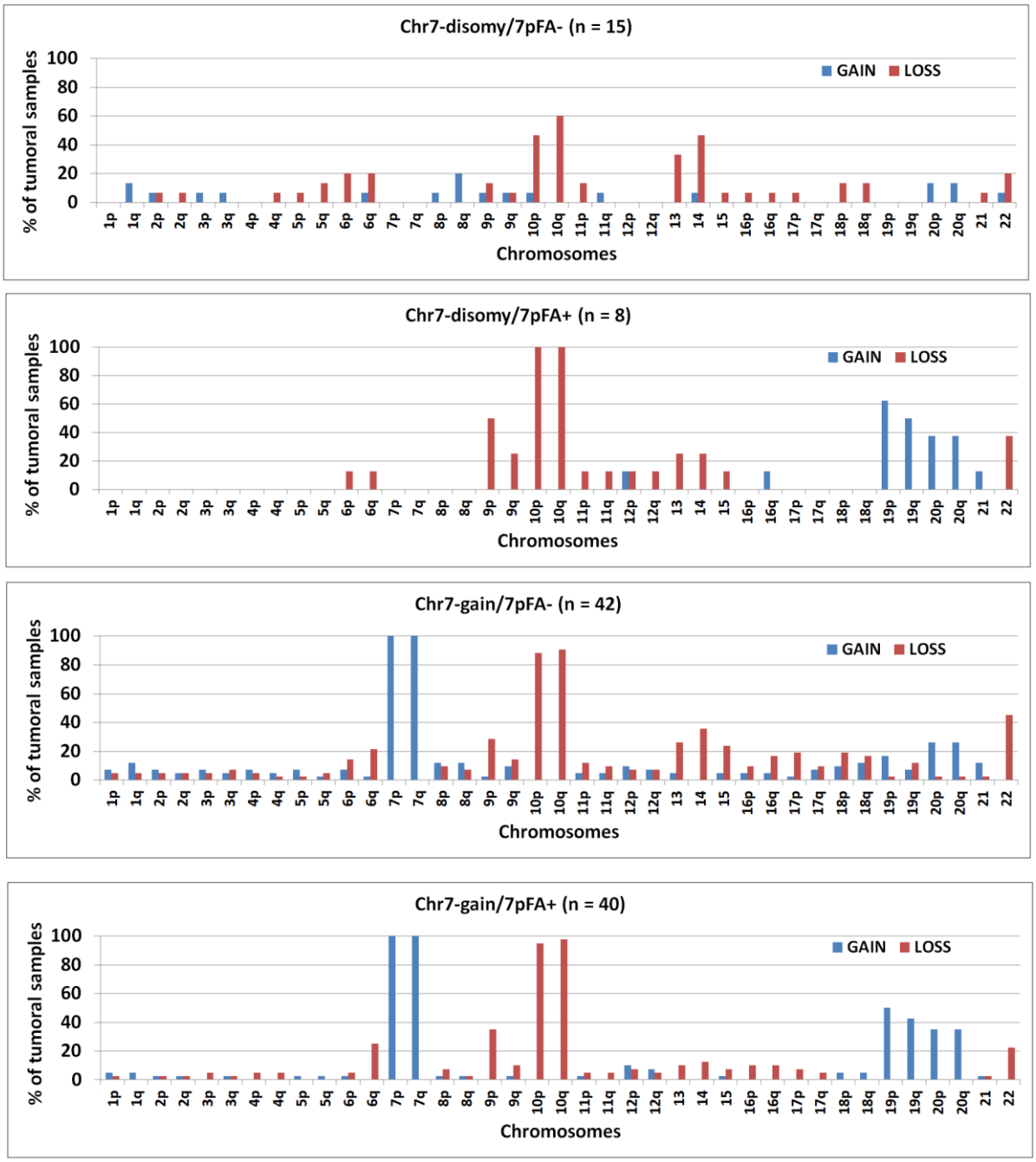


Figure 69. Bar graphs showing the percentage of samples bearing arm-level gains or losses in chromosomes 1-22 in 4 different GBM subgroups according to the presence of chr7gain and/or 7pFA. Only GBM samples profiled for gene expression by RNASeq have been used for this analysis. The number of samples (n) in each group is reported in the corresponding graph

A simple inspection of graphs reported in **Figure. 69** reveals that the frequency of chr19-gain is higher in 7pFA positive groups (chr7-gain/7pFA+ and chr7-disomy/7pFA+) in comparison to 7pFA negative groups (chr7-gain/7pFA- and chr7-disomy/7pFA-). However, the previous groups were formed only with samples analysed by both RNA-Seq and SNP array, because they have been prepared for the following transcriptome analysis.

Table 30 re-proposed the relative risk association of chr7-gain, chr10-loss, chr19-gain and chr20-gain with focal amplification and between chr10-loss and gain of chromosome 7, but for the small cohorts concerning n= 164 samples of Brennan et.al 2013 classification. This analysis confirmed the significant association between chr19-gain and 7pFA(p-value < 0.0001).

	Supplemental Table S5. EGFR RNA Seq Summary Brennan 2013 (n = 164)			
Only samples with SNP array and RNAseq data (n = 128)	7p-FA+	7pFA-	X^2 P-value	Relative Risk (C.I. 95%)
Chr7-gain (n = 107)	40	42	0.2337	1.40 (0.82-2.68)
Chr7-disomy (n = 25)	8	15		
Chr10-loss (n = 104)	53	49	0.0078	∞ (1.44-∞)
Chr10-disomy (n = 7)	0	7		
Chr19-gain (n = 26)	23	3	<0.0001	2.61 (1.84-3.75)
Chr19-disomy (n = 74)	25	49		
Chr20-gain (n = 34)	21	13	0.0477	1.48 (1.00-2.11)
Chr20-disomy (n = 85)	35	49		
	Chr7gain / 7pFA-	Chr7disomy / 7pFA-		
Chr10-loss (n = 49)	37	9	0.0313	1.87 (1.03-5.12)
Chr10-disomy (n = 7)	3	4		

Table 30 Chi-square association test for different chromosomal aberrations and 7pFA. SNP array of (Brennan et al., 2013) classification

4.3.3 GBM classification

The transcriptomics data were retrieved from online resource of GDC data portal. From the total available RNA-seq data (174 samples), only 128 primary tumours and 5 normal brain tissue data were considered. Among primary tumours, 84 samples bearing 7-gain (65% of the total) and 25 samples with 7-disomy (30% of the total). According to Brennan et al. 2018 classification the 7-gain and the 7-disomy samples were first subdivided **1) Classical:** 7-gain (n = 18) and 7-disomy (n=6) **2)Mesenchymal**: 7-gain (n = 29) and 7-disomy (n=6) **3) Neural**: 7-gain (n = 16) and 7-disomy (n=3) **4) ProNeural**: 7-gain (n = 18) and 7-disomy (n=6) **5(G-CIMP:** 7-gain (n = 2) and 7-disomy (n=5) (Brennan et al., 2013).

The 7-gain and the 7-disomy samples were further subclassified in accordance with focal amplification in the EGFR cytoband 7p11.2 (7p-FA+ and 7p-FA-). Four cytogenetics groups were obtained **1) Chr7-gain/ 7p-FA+** (n = 40), **2) Chr7-gain / 7p-FA-** (n=42), **3) Chr-7disomy/7pFA-** (n=15) and **4) Chr7-disomy /7pFA+** (n=8).

In **figure 70**, the pie charts show the percentages of the GBM subtypes for each analysed group. The first observation concerns the highest percentages of classical subtype with focal

amplification (7pFA+) in those groups bearing chr7-disomy (62%) and chr7-gain (40%). In Chr7-disomy /7pFA- the most representative subtype is proneural (33%) followed by classical (27%) and C-cimp. Mesenchymal subtypes characterized Chr7-gain /7pFA- group for 41%, followed by proneural subtype for 31 %. In Chr7-gain and chr7-disomy without EGFR status classification there is not a clear emerging subtype.

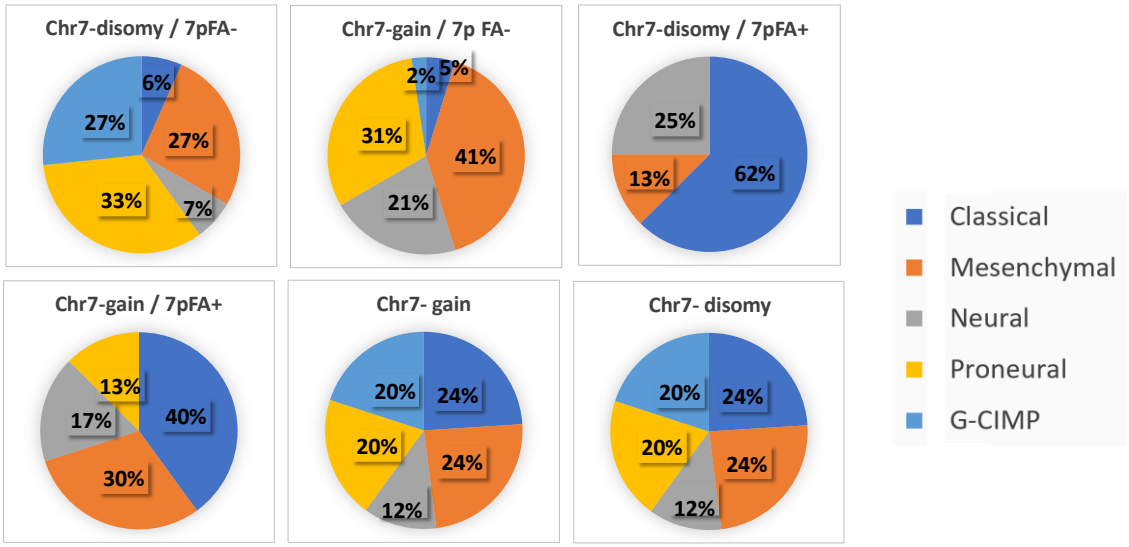


Figure 70. Pie-Charts showing subclassification in of subtype Chr7-gain/ 7p-FA+ (n = 40), 2) Chr7-gain / 7p-FA- (n=42), 3) Chr-7disomy/7pFA- (n=15) and 4) Chr7-disomy /7pFA+ (n=8) and the general subgroups Chr7-gain and Chr7-disomy.

4.3.4 Transcriptomics analysis GBM classification

A differential expression analysis of transcript levels was performed on the selected groups bearing the chromosomal aberration of 7-gain or not bearing 7-disomy [Chr7-gain/ 7p-FA+ (n = 40), Chr7-gain / 7p-FA- (n=42), Chr7-disomy /7pFA+ (n=8)] in comparison with the control group Chr-7disomy/7pFA- (n=15). The analysis was performed by edgeR package (McCarthy et al., 2012; Robinson et al., 2010) and expressed as Fold-Change (FC). We called OverT (Overexpressed Transcripts) those transcripts expressing a value of FCvsCTRL > 1.3 at a False Discovery Rate adjusted p-value (adjp) <0.05. The chromosomal distribution of OverT enriched genes are expressed as normalized chromosomal distribution index (NCDI); (Condorelli et al., 2019). The analysis results are reported in **Figure 71**.

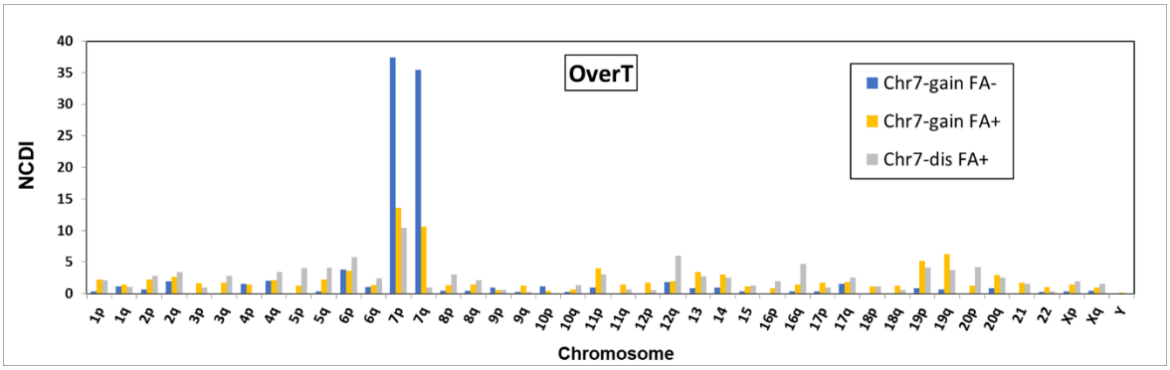


Figure 71: The normalized chromosomal distribution index (NCDI) of OverT of each cytogenetic group bearing (Chr7-gain) or not bearing (Chr7-dis) the gain of chromosome 7 with (7p-FA+) or without (7p-FA-) vs control group (diploid status of chromosome 7).

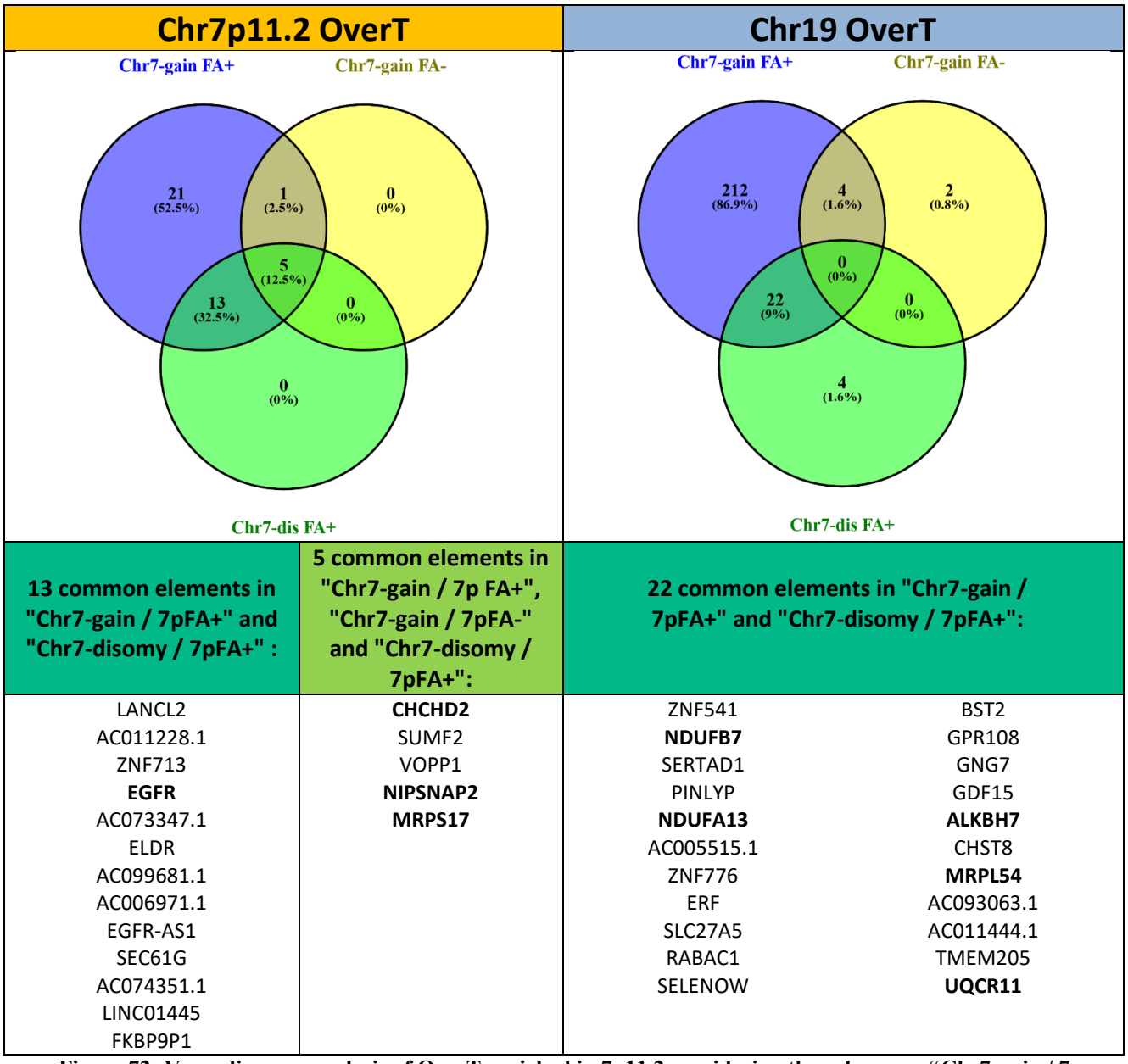
The enrichment of transcripts class in Chr7-gain/ 7p-FA+ group is showed in **Figure 71.** OverT are mainly located on 7p chromosomal arm (38%) and for in 7q chromosomal arm (34%). From these results we could assume that gain of chromosome 7 is itself a sufficient cause to produce the tumour phenotype in GBM. The co-occurring of chromosome 7 gain and 7p-FA+ could reinforce the effects produced by EGFR focal amplificationon pathophysiology of GBM. The number of enriched OverT classes in the analysed groups are showed in **table 31**.

OverT genes			
	Chr7-disomy / 7pFA+	**Chr7-gain / 7pFA+**	**Chr7-gain / 7pFA-**
All chrs	334	2043	405
Chr7	28	467	285
Chr7p11.2	18	40	6
Chr19	26	238	6
***FCvsCTRL**>1.3, adjp<0.05;			

Table 31 enriched OverT classes in the analysed groups

In agreement with transcriptional gene-dosage effects the highest number of OverT is observed in chr7-gain/7pFA+ group (467 genes), followed by chr7-gain/7pFA- (285 genes) and chr7-disomy/7pFA+ (only 28 OverT genes). However, when genes localized in chromosomal band 7p11.2 were taken into consideration (7p11.2-OverT genes), only 6 genes were detected in chr7-gain/7pFA-, while Chr7-gain/7pFA+ and chr7-disomy/7pFA+ showed 40 and 18 genes, respectively. Since chromosome alterations involving whole chromosome gain or focal amplifications are likely to act through a modification of transcriptional activity of genes located in the aberrant region (gene-dosage transcriptional effect), we determined the OverT genes shared among the three cytogenetic groups bearing chr7-gain and/or 7pFA. Chromosome 19-gain is significantly associated with 7pFA (**Table 29-30**), we identified the OverT genes located on chr19 (chr19-OverT) and shared between the two groups bearing 7pFA. As shown in **Table 31**, 26 shared chr19-OverT genes were so identified.

A venn-diagram analysis was performed to individuate which genes are shared between the groups indicated in **Table31**.



13 common elements in "Chr7-gain / 7pFA+" and "Chr7-disomy / 7pFA+" :	5 common elements in "Chr7-gain / 7p FA+", "Chr7-gain / 7pFA-" and "Chr7-disomy / 7pFA+":	22 common elements in "Chr7-gain / 7pFA+" and "Chr7-disomy / 7pFA+":	
LANCL2	**CHCHD2**	ZNF541	BST2
AC011228.1	SUMF2	**NDUFB7**	GPR108
ZNF713	VOPP1	SERTAD1	GNG7
EGFR	**NIPSNAP2**	PINLYP	GDF15
AC073347.1	**MRPS17**	**NDUFA13**	**ALKBH7**
ELDR		AC005515.1	CHST8
AC099681.1		ZNF776	**MRPL54**
AC006971.1		ERF	AC093063.1
EGFR-AS1		SLC27A5	AC011444.1
SEC61G		RABAC1	TMEM205
AC074351.1		SELENOW	**UQCR11**
LINC01445			
FKBP9P1			

Figure 72: Venn-diagram analysis of OverT enriched in 7p11.2 considering the subgroups "Chr7-gain / 7p FA+", "Chr7-gain / 7pFA-" and "Chr7-disomy / 7pFA+" (left panel), and genes located on chromosome 19 for those selected groups.

As shown in **Figure 72**, five 7p11.2-OverT genes are shared among all the three cytogenetic groups (CHCHD2, SUMF2, VOPP1, NIPSNAP2, MRPS17), while another thirteen genes are only shared between the two groups bearing the 7pFA. These results suggest that the transcriptional expression profiles of genes located in 7p11.2 are significantly different between focal amplification and whole chromosome gain.

4.3.5 Gene Ontologies (GO) analysis

In order to understand the weight given from OverT genes to promoting GBM phenotype, a gene ontology analysis was performed. The analysis has considered those genes shared between chromosome 7p11.2 and chromosome 19 as showed in **Figure 72**. Genes were classified by PANTHER database (access May 5, 2020) in cell component (CC) and biological processes(BP). The .json files were analysed by R packages and GoPlot tools. The aim of this method was to figure out at which level the cooperation of genes located on chr7-chr19 intervenes.

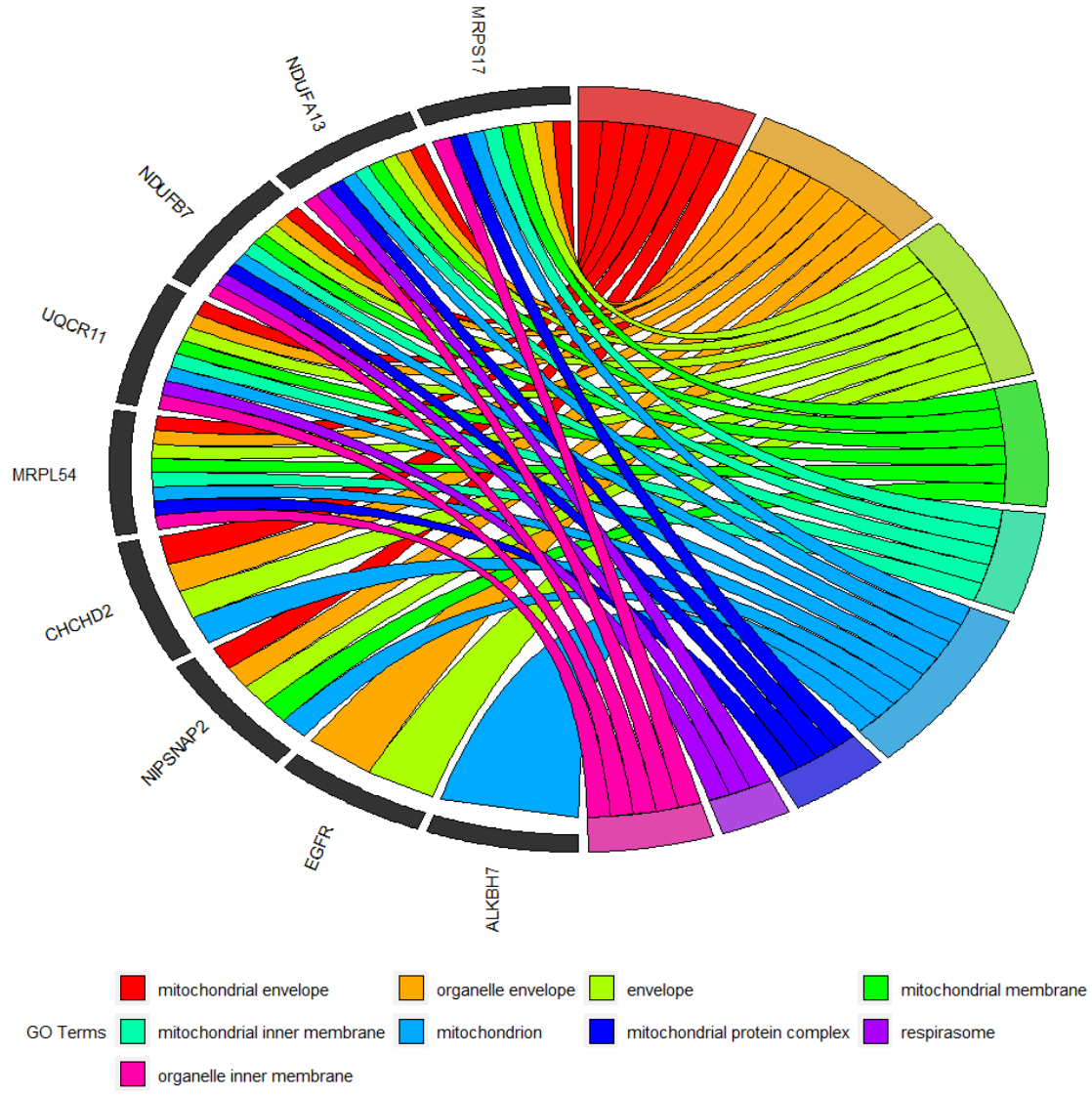


Figure 73: GoPlot and Cell component analysis of common genes between chr7 and chr19.

The **Figure 73** shows the cell component gene ontology. Each coloured branch links one or more genes to the same go term. Mitochondria and respirasome component could be altered in GBM. In particular, CHCHD2 is located on 7p11.2 locus and belongs to a class of proteins

characterized by four cysteine residues, separated by 10 amino acids from one to another. NIPSNAP2 (GBAS), located on 7p11.2 locus, is a mitochondrial matrix protein. It forms a complex with NIPSNAP1

NDUFA13 (GRIM-19), located on 19p13.26, is a cell death regulatory gene that promotes apoptosis and a negative regulator of cell growth, and it is also involved in mitochondrial metabolism. The major role of GRIM-19 in the control of cell growth is exerted through STAT3, a transcription factor inhibited by GRIM-19 binding. GRIM-19 has a unique role in the preservation of mitochondria, in addition to its electron transfer activity (Jin et al.,2010; Lu and Cao, 2008). NDUFB7, located on 19p13.12 locus, is a subunit of the complex I of the respiratory chain. It has separate phosphorylation sites for both protein kinase A and PDH kinase. Phosphorylation by PKA induces higher levels of electron flow and increases ATP synthesis whilst decreasing ROS formation. On the contrary, phosphorylation with PDH kinase decreases the enzyme activity, possibly in condition of starvation, and augments ROS production (Raha et al., 2002).

4.3.6 Hierarchical clustering of brain tumours samples

In **figure 74**, we performed a hierarchical clustering analysis by selecting these gene located on chr7p11.2. The TPM data were pre-filtered by excluding all transcripts with 0 value in more than 70% of samples. The tpm values were normalized against the average of control group. As show in the legend in **Figure 74** the data were further clustered in accordance with 7pFA status (red = FA+; blue = FA-) as well as the cytogenetics groups **1) Chr7-gain/ 7p-FA+** (n = 40) (orange), **2) Chr7-disomy /7pFA+** (n=8) (fuchsia), **3) Chr7-gain / 7p-FA-** (n=42) (green), **4) Chr-7disomy/7pFA-** (n=15) (dark green); 105 samples were analysed in total. Data are clustered by the unweighted pair group method with arithmetic mean (UPGMA) and Euclidean distance.

The gene expression profile in cytogenetic band 7p11.2 determined a good clustering of GBM samples according to the presence or absence of 7p-FA, independently from the presence of chr7-gain. Such results can be explained by several factors: 1) the copy number changes are quantitatively limited in case of chr7-gain (trisomy or tetrasomy of the chromosome), while focal amplification can occur in >4 copies and each gene can show a non-linear response to copy number variations; 2) the gene transcriptional changes associated to 7pFA can trigger downstream effects or promote evolutionary pathways that modify the expression profile of other genes located in band 7p11.2.

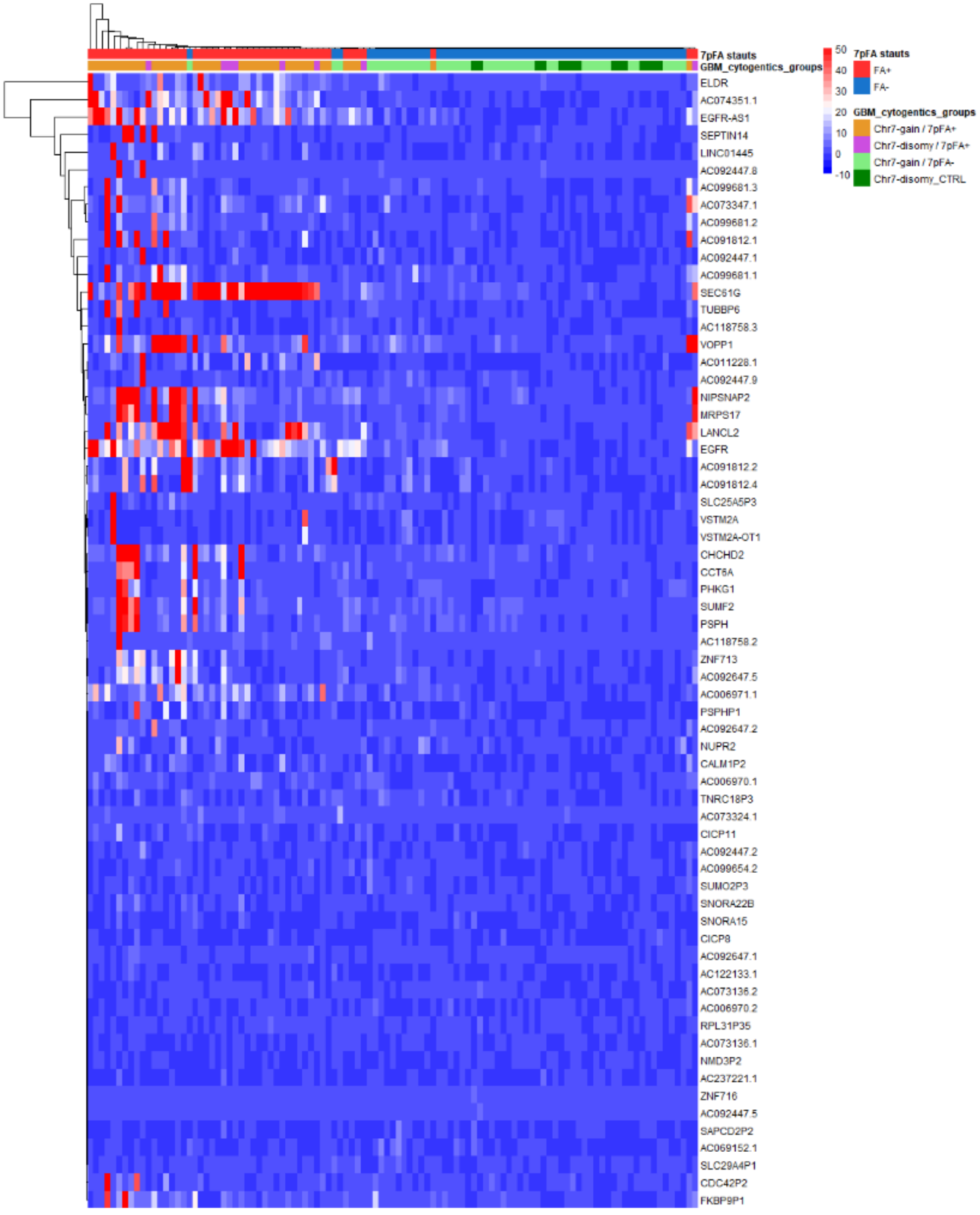


Figure 74: **Hierarchical clustering of 66 genes located on chromosomal region of 7p11.2 Data are clustered by the unweighted pair group method with arithmetic mean (UPGMA) with Euclidean distance.**

To better investigate the genes located on chr7p11.2 in the analysed cytogenetics groups. The **Figure 75** shows the tpms values for **Chr7-gain/ 7p-FA+** (n = 40), **2) Chr7-gain / 7p-FA-** (n=42), **3) Chr-7disomy/7pFA-** (n=15), **4) Chr7-disomy /7pFA+** (n=8) and 5) **Normal brain tissue** (n=5). The table at the bottom of Figure 75 indicates (through a red asterisk) if that genes belong to OverT class, in other word if that gene is over-expressed respect to the control group not bearing any 7-gain and 7pFA **(Figure 75)**.

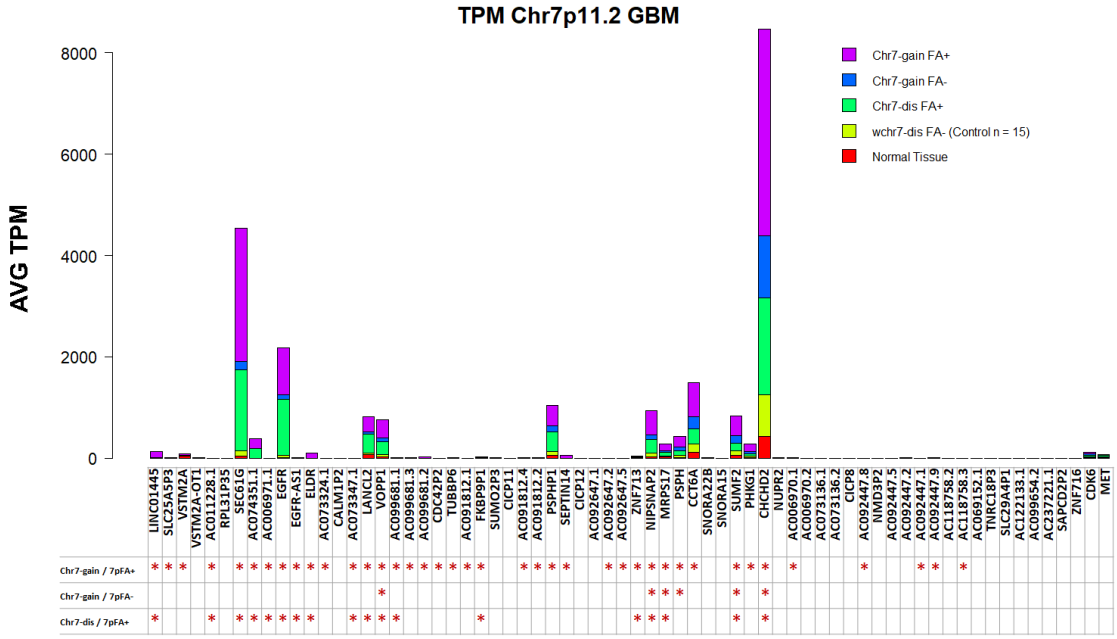


Figure 75: Transcripts per million (TPM) of genes located on chr7p.11.2.

High expressed genes are enriched in those groups with 7p-FA+ (gain or disomy). Examples are SEC16G, EGFR, LANCL2, VOPP1, NIPSNAP2, CHCHD2. The genes linked to the focal amplification and the gain of chromosome 7 are SEPTIN14, CCT6A.

4.3.7 Gene Set Enrichment Analysis (GSEA) of OverT genes.

As reported in the previous sections, the identification of transcriptionally dysregulated genes leads to recognise which pathways are altered in GBM tumorigenesis. Several transcriptionally dysregulated genes could be considered only passenger modifications that do not participate in a significant way to the generation of the cancer phenotype. In order to generate hypothesis on the functional role of transcriptionally dysregulated genes located in chr7 we performed a gene set enrichment analysis.

Three cytogenetic groups bearing chr7-gain and/or 7pFA were compared with the CTRL group (chr7-disomy/7pFA-). TPM values are used. In order to force the analysis towards the discovery of functional interactions among genes located on chr7 and 19 we performed an

analysis using as input data only genes located on those chromosomes. Interestingly, the top-ten pathways revealed by such analysis in the group chr7-gain/7Pfa+ were also among the top-ranked position in the other two cytogenetic groups (**Figure 76**).

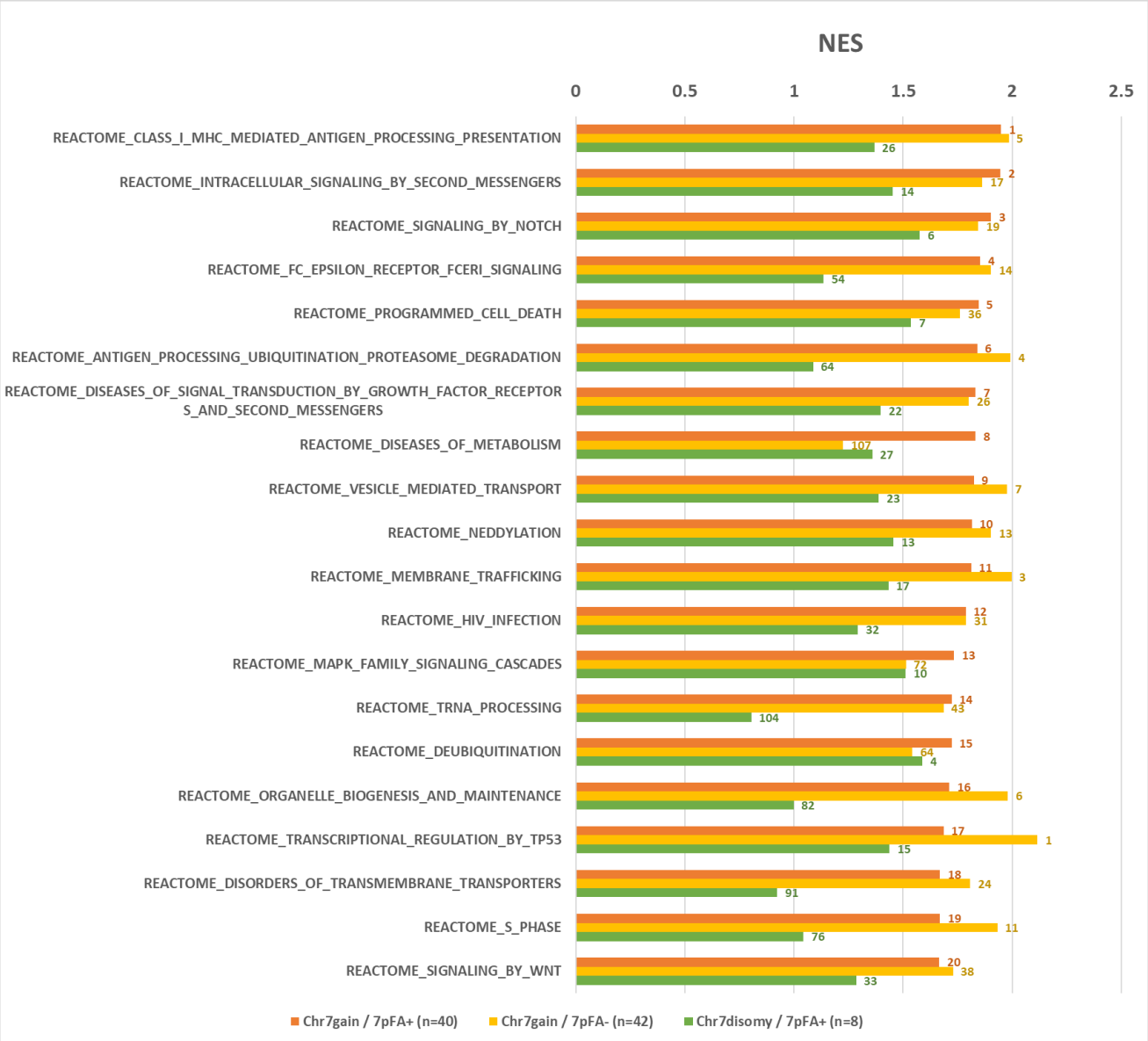


Figure 76: GSEA analysis. Top ten pathways in Chr7-gain/ 7p-FA+. NES decreasing values. The number at the end of each histograms indicated the position of the same pathway in the GSEA analysis for Chr7-gain / 7p-FA- (n=42) and Chr7-disomy /7pFA+.

The **figure 76** is an integrative summary of the three GSEA analysis performed for Chr7-gain/ 7p-FA+ (n = 40), 2) Chr7-gain / 7p-FA- (n=42), 3) Chr7-disomy /7pFA+ (n=8) against the 4) Chr-7disomy/7pFA- (n=15). In figure 76 are only showed the ten more representative pathways for Chr7-gain/ 7p-FA+. These pathways are ordered by Normalized Enrichment Score (NES) (see the orange histograms). The individual GSEA analyses concerning Chr7-gain / 7p-FA- and Chr7-disomy /7pFA+ are flanked in the top-ten pathways of Chr7-gain/ 7p-FA+. A number at the end of each histograms indicates the relative pathways position. It is

relevant the "Reactome_Programmed_cell death" pathways because Malignant glioma undergo necrosis rather than apoptosis when there is an energy deprivation (Steinbach et al., 2003). "Reactome_transcriptional_regulation_by_tp53" is at the first position in Chr7-gain / 7p-FA- (n=42).

4.3.8 Recurrent Point Mutations in GBM

Recurrent point mutations by WES data are shown in oncoplots in **Figures 77**. The point mutation data are only 60 samples because they are the only having id_cases matchable with cytogenetic, rnaseq and wes data from TCGA. In total we found 2926 SNPs in 2392 genes. 1537 Silent variants were excluded. The main subtypes (Classical, Mesenchymal, Neural, Pro-neural and C-cimp) and the many cytogenetics groups are indicated by the annotation bar below the graph. High frequency of TP53 mutations (mainly missense mutations) are identified in **Chr7-gain/ 7p-FA+ group** and **Chr7-gain / 7p-FA- group.** Instead, EGFR characterizes the Chr7-gain FA+ Group, while it is not present in the FA-. TP53 mutations are inversely proportional to EGFR mutations in the chr7-gain FA-. Classical subtype is characteristic of Chr7-gain FA+ Group while Proneural subtype of Chr7-gain FA-.

Figure 77: Oncoplot showing point mutations detected by WES in 60 samples analysed in TCGA study.

4.3.8 GBM cytogenetic groups and prognosis.

We evaluated the prognostic impact of broad gain and focal amplification on chromosome 7, organizing the entire cohort of TCGA GBM samples into four groups: chr7-disomy FA- (n=55), chr7-gain FA+ (n=159), chr7-gain FA- (n=194), chr7-disomy FA+ (n=21). Chr7-disomy FA- showed a better overall survival (OS, median survival 22.48 months) in comparison to the other three groups (14.20, 12.69, 9.13 months, respectively; **Figure. 78**). Similar results were obtained for progression free survival (PFS, **Figure 78**). No significant difference was observed among chr7-gain FA- (n=194), chr7-disomy FA+ (n=21), chr7-disomy FA- (n=55). Therefore, both the presence of FA on chr7 or a broad gain of the same chromosome are associated to a worsening of the prognosis.

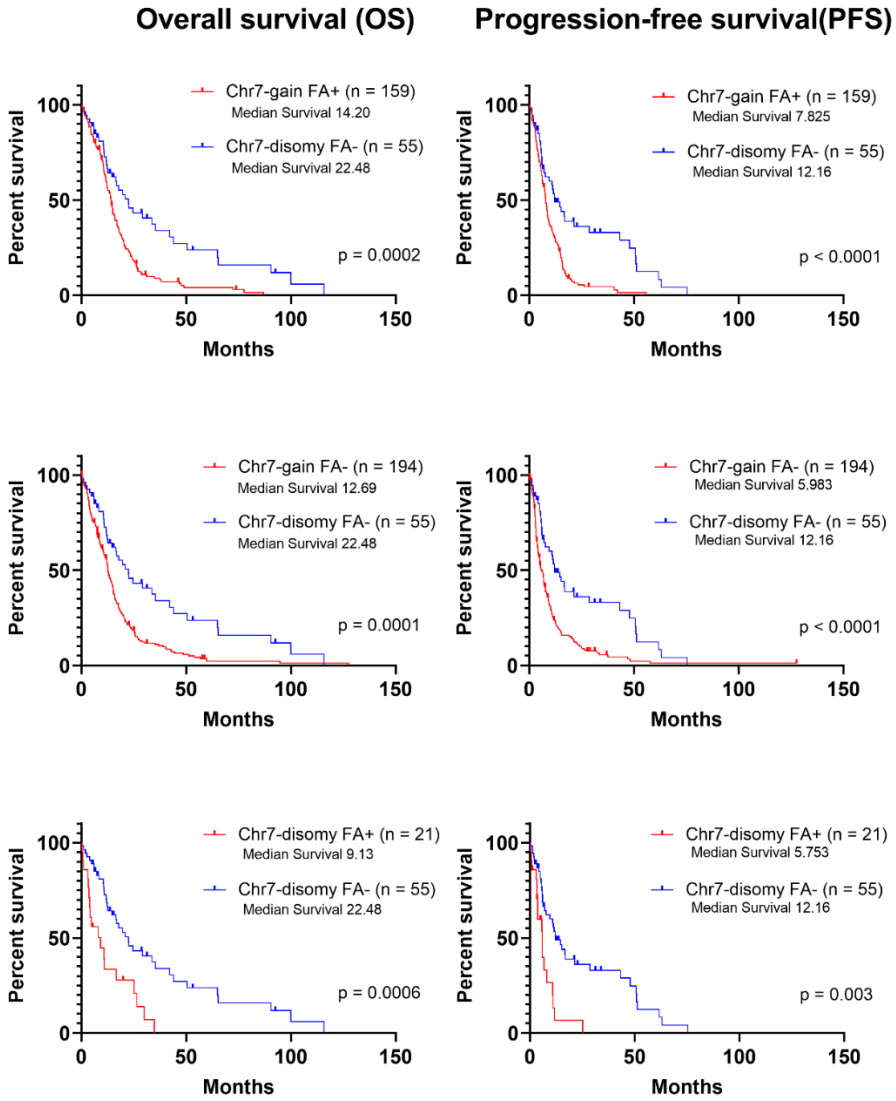


Figure 78: Kaplan-Meier curves comparing Overall Survival (OS, Left graphs) and Progression Free Survival (PFS, right graphs) between Chr7-disomy FA- samples and the other three subgroups (chr7-gain FA+, top; chr7-gain FA-, middle; chr7-disomy FA+, bottom). Log-rank (Mantel-Cox) test was used for statistical significance and corresponding p values are reported in each graph.

5. DISCUSSION AND CONCLUSION

5.1 BREAST INVASIVE CARCINOMA (BRCA)

In our work, we focused on a specific subpopulation of breast cancers (with 1q-gain and/or16q-loss), and our cytogenomics classification was a tool to classify BRCA adenocarcinomas in different "chromogroups" according to the presence of different combinations of Chr 1 and 16 copy number abnormalities and to investigate the associated dysregulated genes. These combinations were designed to correspond to the expected patterns of copy number abnormalities generated by chromosomal aberrations previously found in classical cytogenetic studies of breast cancer, such as der(1;16) (q10;p10), i(1q), and del(16q). Indeed, the association of 1q-gain and 16q-loss can be produced by the single chromosomal aberration, such as der(1;16), or by the combination of two different aberrations, such as i(1q) and del(16q). The analysis of aneuploidy scores revealed an enrichment of samples with low values (median 6) in group A, whose copy number criteria were inspired by der(1;16), or in subgroup D1, inspired by del(16q), in agreement with the observation that those aberrations are often observed as the sole cytogenetic anomalies in breast cancer (Dutrillaux et al., 1990; Farabegoli et al., 2004; Kokalj-Vokac et al., 1993; Muthuswami et al., 2013; N. Pandis et al., 1994; Nikos Pandis et al., 1992, 1995; Rye et al., 2015; Teixeira et al., 2002; Tsarouha et al., 1999). Higher aneuploidy scores were observed in samples with 1q-gain without aberrations of chr16 (groups C or B2: median scores of 9.5 and 14, respectively) or with concomitant 1q-gain and 16q-loss determined by two different co-occurring cytogenetic abnormalities such as i(1q) and del(16q) (subgroup B1: median score of 21). The latter observation suggested that a more complex evolutionary process, based on chromosomal instability, is involved in the progression of cancers of subgroup B1. Nonetheless, the relatively high frequency of tumours potentially bearing the der(1;16) (group A included 36% of all tumours showing 1q-gain or 16q-loss) and the fact that those tumours bear a low number of other chromosomal aberrations (low aneuploidy score) supported the hypothesis that genes located in 1q and 16q might play a strong cooperative cancer driver effect. In early cytogenetic studies in breast cancer, high pathogenic impact of 1q-gain was inferred from the observation that the two most frequent chromosomal aberrations, der(1;16) and i(1q), have 1q in common (Dutrillaux et al., 1990; Nikos Pandis et al., 1992; Teixeira et al., 2015) However, following studies supported the importance of 16q-loss in the absence of chr1 aberrations in breast cancer (Bürger et al., 2013;

Farabegoli et al., 2004; Roylance et al., 2006).The data analysis performed in the present thesis confirmed the relatively high frequency of 16q-loss without chr1 aberrations and with a low aneuploidy score (group D included 15% of all tumours showing 1q-gain or 16q-loss). We reported that the CDH1 differential expression between ductal and lobular carcinomas was maintained in all examined 1,16 chromogroups that included a significant number of both histotypes (**Figure 28**). Collectively, these data indicated that transcriptional decrease of CDH1 is only a weak effect in invasive ductal carcinomas, suggesting that the frequent loss of 16q in such histological subtype might be explained by the cancer evolutionary advantage due to reduced transcription of other 16q genes (Hungermann et al., 2011) or by a higher sensitivity to CDH1 haploinsufficiency due to the cooperative effect of other gene mutations. In light of the frequent association of 1q-gain and 16q-loss in both invasive and lobular breast carcinomas, we reasoned that investigation of cooperative functional links between transcripts encoded in those chromosomal arms was a valid strategy for the identification of novel candidate driver genes underlying the selection of those recurrent chromosome aberrations. Previous studies, aimed to identify driver genes in 1q and 16q, separately analysed expression of genes located in these chromosomes (Hungermann et al., 2011; Muthuswami et al., 2013). Indeed, the chromogroup A, enriched in der(1;16), was characterized by a low aneuploidy score, thus suggesting that the putative interchromosomal 1q/16q gene cooperation can be an early event and is not associated with an extensive chromosomal instability. Such cooperation might involve the early E-cadherin loss in lobular carcinomas (McCart Reed AE, Kutasovic JR, Lakhani SR, 2015). On the contrary, E-cadherin loss occurs as a late event in invasive ductal carcinomas (Alsaleem et al., 2019; Jeschke et al., 2007), and the understanding of the functional meaning of the early and frequent 1q-gain/16q-loss co-occurrence requires the definition of further mechanisms. In order to select putative cancer driver genes located in chr1q and chr16q, we exploited comparisons of the corresponding transcript levels between chromogroups bearing 1q-gain and/or 16q-loss and a so-called “CTRL group,” i.e., a cancer group devoid of any arm-level aberrations of chr1 and chr16. The main assumption for this strategy was that the analysis of differential expression between those cancer groups could identify transcriptionally dysregulated genes sensitive to gene dosage effect and that this subset of genes is enriched in cancer driver genes. With this in mind, we investigated functional cooperation between genes located in 1q and 16q by two different methods. In the first one, the selection of candidate genes was based on pre-established thresholds (linear FC and adjp values) in the comparisons between the selected 1,16-chromogroup and the CTRL group or between the selected chromogroup and the normal breast tissue group. Moreover, the

concordance of transcript level changes with corresponding copy-number abnormalities in the different 1,16-chromogroups was taken into account. Based on previous data on aneuploidy-induced transcriptional changes (Condorelli et al., 2019; Condorelli et al., 2018), we selected OverUpT genes located in 1q and UnderT genes located in 16q and submitted those gene lists to the over-representation analysis of functional pathways (Zhou et al., 2019). In the second strategy, differential gene expression between different chromogroups and the "CTRL cancer group" was investigated by gene set enrichment analysis (Subramanian et al., 2005). Both methods showed concordant results, pointing out the involvement of functional pathways that show the cooperation of genes located on 1q and 16q, such as "NOTCH2 Activation and Transmission of Signal to the Nucleus," "NOTCH3 Activation and Transmission of Signal to the Nucleus," and "Formation of the beta-catenin:TCF transactivating complex." Indeed, the involvement of Notch signaling system in breast cancer has been repeatedly suggested in the literature (Mollen et al., 2018).

The overall survival analysis performed comparing the two main cohorts (TCGA and MEATBRIC studies) of clinical data in breast cancer has allowed to categorize the 1,16 chromogroups and the intrinsic subgroups in a survival rate scale. Focusing on TCGA study, the median survival (MS) of the 1,16 chromogroups show a more favourable prognosis. In particular, Group C (n= 90, MS = Undefined), Group A (n = 178 and MS = 212.2 moths) have the favourable median survival followed by Group D (n = 75 and MS = 127.3 months) to Control (n = 71 and MS = 113,8) Group B (n=171 and MS = 102.8 months) (**Figure 33**). However, the TCGA prognosis data do not allow a defined interpretation due to the short follow-up available period and the small number of patients. The METABRIC analysis has re-organized the 1,16 chromogroups groups with the following decreasing survival scale: Control (n=292 and MS =199 months), Group D (n= 183 and MS =184.7 months), Group A (n= 447 and MS =180.8 months), Group C (n= 158 and MS = 128.5 months) and Group B (n= 154 and MS = 112.6 months). The control groups not bearing any chromosomal aberrations shows the best median of survival. The Group B is the worst groups in both cohorts. Concerning the intrinsic molecular subtypes Luminal A (n=700 and MS = 186.6 months) shows the best prognosis survival followed by Normal (n=148 and MS = 159.2 months), Basal (n= 209 and MS = 130.9 months) Luminal B (n=475 and MS = 123 months) and Her2 (n=224 and MS = 106.6 months). Using the portion of group A (better prognosis) and group B (worse prognosis) a comparison with LumA and LumB was performed (**Figure 36**). GroupA_LumA (n = 278 and MS = 197.8), Group B_ LumA (n= 36 and MS = 172.8), Group A_ LumB (n = 88 and Ms = 139.6) and the Group B_LumB (n = 67 and MS = 118.1). The

results open the hypothesis to associate an alternative method to assess prognosis with cytogenetic classification of 1,16 chromogroups.

The secretory carcinomas show particular gene feature, the ETV6-NTRK3 fusion gene (Pareja et al., 2021). In the study of Laé et al. secretory breast carcinoma, classified as basal like, shows a favourable prognosis (Laé et al., 2009). Although secretory breast carcinomas have a favourable outcome, there is not a clear method to distinguish them from the other breast cancer types as well as to diversify the therapeutic approach. We proposed a new approach through transcriptomics analysis to discriminate tumour samples bearing the gene feature ETV6-NTRK3 by using the discriminatory threshold of 7.38 TPM (calculated by summing the median plus 2-fold the standard deviation). The NTRK3+ have a better prognosis compared to the entire TCGA cohort, but they show a favourable prognosis also in the context of basal subtype (Log-Rank test significance 0.0169). The data confirm the subtype of secretory cancers even though broader cohorts could clarify our method. Unfortunately, the METABRIC cohorts is not suitable for this purpose because the microarrays technologies used in this study do not well discriminate the NTRK3+ from NTRK3-. Future perspective could involve broad cohorts analysed by single cell RNA-seq.

The experimental approach based on investigation of similar gene profiling of 1q over-expressed genes between transcriptomics cell data and the tumours samples led us to the identification and the usage of the suitable cell model bearing (CAL148) and not bearing (CAL51) the main chromosomal aberration of der(1,16) in breast cancer. The cell growing and maintenance allowed a better understanding of behaviours and peculiarities of CAL148 and CAL51. CAL148 has showed a clear response to the EGF treatment (**Figure 42**) at 24h, 48h, 72h, 144h timepoints and at four cell density (2000 cell/wells, 4000 cell/wells, 8000 cell/well and 16000 cell/wells). The differences in all conditions were statistically significant and tested by un-paired t test. The greatest EGF effect is observed between 96h and 144h hours both at 4000 and 8000 cells/well. CAL51 cells show a high growth rate and a resistant phenotype at EGF treatment neither in the phase of initial adhesion (EGF treatment at plating) and in the following cell growth (EGF treatment 24h after plating) (**Figure 43**). Moreover, work in this thesis highlights a possible mis-classification of CAL148 as triple-negative breast cancer (TBNC). In fact, usually CAL148 are considered TBNC because of negative response to immunohistochemistry assays concerning estrogen receptors (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2/ERBB2) antibodies. Re-testing the HER2 receptor in immunocytochemistry(ICC) assays (**Figure 53A-B**) and analysing the transcriptomics data (RNA-seq) (**Table 21**) allowed us to verify an expression of HER2 both

in ICC assay and RNAseq data (200.966 TPM). On the bases of these evidence questions to which we would like to find an answer are: what does CAL148 so responsive to EGF treatments? Is there an influence of HER2/ERBB2? To answer to these questions we analysed the transcripts levels in CAL148 and CAL51 (**see table 21**) finding that in CAL148 EGFR transcripts is not abundant (13.178 tpm) and its expression may suggest that other receptors might be responsible for the effect of EGF treatments. For instance, the epidermal growth factor receptor 3 (ERBB3) (161.43 tpm) and ERBB4 (6.58 tpm) are expressed in CAL148 and it is possible to hypothesize a role of ERBB3 in CAL148 due to its high transcripts expression. Receptor tyrosine kinases (ErbBs) are essential to transduce extracellular signalling. ErbB`s ligands bound the extracellular ligand-binding ectodomain inducing homo- and heterodimerization. The ErbBs dimerization is the key of signal transduction by activation of intracellular tyrosine kinase domain followed by the phosphorylation of the C-terminal tail domain (Linggi et al., 2006). The phosphorylation activates downstream signalling of pathways such as Ras/MAPK or PLCγ1/PKC, as well as PI(3)kinase/Akt, and STAT pathways (Scaltriti et al. 2006). EGFR and ErbB4 are fully independent from dimerization to activate signalling transduction. Among the possible combinations of homo- and heterodimer are known 1) EGFR and ErbB2, 2) EGFR and ErbB3, 3) ErbB3 and ErbB2 and 4) ErbB3 and ErbB4 (Citri et al., 2003; Graus-Porta et al., 1997). In particular, the specific ErbB2 ligands are unknown but its preferred partner of dimerization is EGFR or also ErbB3 (Graus-Porta et al., 1997). The possible ligands of the ErbBs family (Wieduwilt et al, 2008) are summarized in **table 32.** The Differential response to the EGF treatment observed in CAL148 in comparison to CAL51 and MCF7 could be due to a differential expression of the ErbB´s ligands. As it is showed in the **table 32** and **Table 21** and cross-referencing the data between TPM of CAL148 for EGFR, ErbB2, ErbB3 and ErbB4 and the ErbBs´s ligands for cell lines, the most abundant transcripts of EGF receptor in CAL148 are ErBB2 (200.966 TPM), and Erb3 (161.43 TPM) while the only two ligands of EGF receptor family are the Epiregulin (EREG) (15.41 TPM) and Neuroglycan (CSPG5) (7.745753 TPM) in CAL148.

TPM CAL148	13.1781	200.966	161.43	6.589							
	EGFR	ErbB2	ErbB3	ErbB4	Gene symbol	Chr	Description	CAL-148	CAL-51	MCF7	MDA-MB-231
EGF	+				EGF	4	epidermal growth factor	0	9.883572	0	0
TGF-α	+				TGFA	2	Transforming growth factor alpha	0	0	86.98224	3.450203
Amphiregulin	+				AREG	4	amphiregulin	0	3.294524	288.781	106.9563
HB-EGF	+				HBEGF	5	heparin binding EGF like growth factor	0	0	86.98224	0
Betacellulin	+			+	BTC	4	betacellulin	0	0	0	0
Epigen	+			+	EPGN	4	Epigen	0	0	0	0
Epiregulin	+			+	EREG	4	epiregulin	15.49151	0	24.35503	0
Neuregulin-1			+	+	NRG1	8	neuregulin 1	0	0	24.35503	0
Neuregulin-2			+	+	NRG2	5	neuregulin 2	0	0	0	0
Neuregulin-3				+	NRG3	10	neuregulin 3	0	0	0	0
Neuregulin-4				+	NRG4	15	neuregulin 4	0	0	0	0
Tomoregulin				+	TMEFF1	9	transmembrane protein with EGF like and two follistatin like domains 1	0	0	0	0
Neuroglycan C			+		CSPG5	3	chondroitin sulfate proteoglycan 5	7.745753	16.47262	0	0

Table 32: ErBs receptors ligands and theis Transcript per Million (TPM)

These combinations could be responsible of the EGF treatment response we have observed in CAL148.

Concerning the validation of transcript levels of over-expressed genes in group A which have been identified by RNAseq analysis through bioinformatics pipeline, and have been experimental tested through qPCR and cell viability MTT assay, we can affirm that expression trends concerning NCSTN, CDH1 and BCL9 genes confirm the TPM values observed in RNA-seq data in CAL148, CAL51, MCF7 and MDA-MB-231.

The γ-secretase complex is able to cleave various proteins within their transmembrane domains. It is composed by APH1A, NCSTN and PSEN2 proteins and an increased expression of those genes in hormone receptor-positive breast cancers has been previously reported (Filipović et al., 2011; Peltonen et al., 2013). Some subunits of the gamma-secretase complex were found to be encoded by genes located on 1q (APH1A, PSEN2, and NCSTN) and were overexpressed and upregulated in cytogenetic groups bearing 1q-gain (Privitera et al., 2021). APH1A, NCSTN and PSEN2 enter in the NOTCH signalling and its transduction signalling. The γ-secretase (GS) inhibition has been broadly proposed as potential strategy to turn off the dysregulated Notch signalling in cancer. However, the γ-secretase inhibitors (GSI) do not act on Notch receptor equally (Ran et al., 2017). In cell viability MTT assays

conducted in this PhD thesis, the γ-secretase inhibitors PF 3084014 hydrobromide and DAPT were used. In CAL148 treated with PF 3084014, we observed a drastic cell death at 24h, 48h and 72h with the half maximal inhibitory concentration (IC_{50}) close to 10.86 µM in CAL148 EGF+ at 24h. Increasing the time of treatment the IC50 value is reduced at 6.58 µM in CAL148.These results and these preliminary experiments can suggest a real promising therapeutic candidate target as well as a specific altered pathways associated to the specific chromosomal aberration der1,16 borne by CAL148.

To further investigate the involvement of notch pathways and its elements, a siRNA interfering treatment on CAL148 cells was performed. A reduction of the PSEN2 mRNA expression of around 50% at 72h was observed at 40nM siRNA concentration (**Figure 55B**). The transfection on CAL148 by using the combination of two siRNAs (APH1A+PSEN2) and (NCSTN+PSEN2) produces a reduction of cell viability -15.59% and -9.52% respectively.

Future efforts will focus on the silencing of the gamma secretase complex in combination with BCL9, a target gene active in the Wnt enhanceosome and b-catenin formation pathway.

The integrated genomic analysis of 1,16-chromogroups provided the following insights on pathogenesis of invasive breast adenocarcinomas:

Invasive lobular carcinomas could be observed both in the presence or in the absence of 16q-loss, although 16q-loss-associated lobular carcinomas were much more frequently observed (155/201, 77%), as already reported (Ciriello et al., 2015). Interestingly, 16q-disomic lobular carcinomas were found to be a distinct subgroup of cancers characterized by a near-euploid karyotype, suggesting that a different form of genome instability is driving this cancer subtype. In the presence of a significant aneuploidy score (>4), 16q-loss was found to be a main determinant of lobular carcinomas, as shown by the lack or rarity of this phenotype in 16q-disomic groups B2 and C. The frequent co-occurrence of 1q-gain and 16q-loss could be observed in both ductal and lobular carcinomas, although a substantial proportion of lobular carcinomas (group D) could occur in the absence of 1q-gain.Transcriptome and pathway analysis revealed several dysregulated 1q- and 16q- genes that are overexpressed or underexpressed in 1,16-chromogroups in both ductal and lobular cancers and highlighted functional networks that may underlie the breast cancer progression. 1q-located genes, such as BCL9 and gamma-secretase components, might play central roles in such cooperating networks. The usage of appropriate cancer cell models that recapitulate the molecular features observed in breast cancers bearing aberrations of chromosomes 1 and 16, has generated a series of testable hypotheses on actionable functional pathways that can be investigated in such models.

5.2 COLON ADENO-CARCINOMA (COAD)

The effect of BCNGs on the level of transcripts encoded in the same chromosomal region (the gene-dosage transcriptional cis-effect) is well-known, and it has been reported in several published studies (Condorelli et al., 2018). The availability of SNP array and RNA-Seq data from a large number of TCGA colon adenocarcinoma samples characterized for genome instability allowed us to devise better control groups in order to identify changes in expression due to specific BCNGs. Indeed, the selective use of CIN samples allowed the generation of tumour control groups devoid of a specific BCNG, although displaying a similar range of other chromosomal aberrations. Our analysis shows that, in chromosomes selected for the presence of recurrent BCNGs, cancer up-regulated transcripts (PositiveT) are more likely to be affected by gene-dosage transcriptional effects than down-regulated ones (NegativeT) (**Figure 61** and **Figure 62**). This phenomenon is also shown by the depletion of OverT among down-regulated transcripts (NegativeT) in BCNG-chromosomes (**Table 25**).

It is easy to conceive that changes in enhancer activity in down- or up-regulated genes can be determinants of the sensitivity to the gene-dosage transcriptional effect. The availability of genome-wide data on the markers of active chromatin in colorectal cancer (Cohen et al., 2017) allowed us to test if the aberrant enhancer activity is associated with any sensitivity to BCNG-associated gene dosage effects. Indeed, a significant enrichment of BCNG-overexpressed genes among the so called "recurrent gained VEL" was observed in this work, suggesting that an active state of chromatin is one of the determinants of the gene-dosage transcriptional sensitivity. Moreover, a high density of recurrent gained VEL and super-enhancers have been observed in some chromosomes (such as chromosome 20 and 13) that undergo frequently to BCNGs in colon cancer. It is possible that the attainment of a saturation level in enhancer activity in cancer-related genes makes the increase in copy number the only efficient mechanism for a further increase of transcriptional activity.

It has been also reported that genes with a high expression level are more dosage sensitive than low-expressed genes (Fehrmann et al., 2015). Our data confirm this observation but reveal that up-regulated transcripts (PositiveT) are more susceptible to gene-dosage effects than down-regulated ones (NegativeT), independently of transcript levels.

Indeed, when transcripts are divided in different bins according to transcript level the difference in dosage sensitivity (measured as the chromosomal density of overexpressed transcripts in selected BCNG chromosomes) between PositiveT and NegativeT is maintained

in all subsets (**Figure 63**). A similar result has been previously observed by using microarray technology for transcriptome analysis (Condorelli et al., 2018).

However, the increased BCNG-associated transcriptional activity does not necessarily translate in cancer growth advantageous features. Indeed, several downstream compensatory mechanisms are present, or the increased transcriptional activity is just a passenger feature that does not affect the functional pathways relevant for cancer progression. In agreement with a functional role in cancer progression, we have previously shown a significant enrichment of cancer-related genes among overexpressed and up-regulated genes (Condorelli et al., 2018). In the present study we show that genes required for cell growth or viability in colon cancer cell lines, identified by genome-scale CRISPR-Cas9 screens (Behan et al., 2019), and called fitness genes, are significantly enriched among BCNG-overexpressed genes. Indeed, the overexpression of specific "cancer fitness" genes might represent an important step in the mechanism underlying the role of recurrent BCNGs in cancer evolution. Hart et al. report some examples of fitness genes that have been found to be amplified across several cancer tissues and cell lines (Hart et al., 2015).

Our data on the relationship between colon cancer fitness genes and recurrent BCNGs are in agreement with recent data obtained by a gain-of-function screening with a modular ORF expression system (Sack et al., 2018). This screening effort is, somewhat, complementary to CRISPR-Cas9 screening inactivating proposed by Behan et al. (Behan et al., 2019). Sack et al. have defined experimentally overexpressed genes that significantly enhance cancer proliferation as GO genes (Sack et al., 2018). They show that taking in consideration GO genes significantly improved the predictions of the frequency of deletions and amplifications of chromosomal arms and whole chromosomes within a pan-cancer tumour set.

Recognition of overexpressed cancer fitness genes in selected chromosomes undergoing recurrent BCNGs, such as Chr20, 8, 13 and 7, allowed a combined pathway analysis of this gene list and the identification of several significant canonical pathways linked to cancer progression. Indeed, intra-chromosomal and inter-chromosomal cooperation of multiple genes can explain the recurrence of BCNGs and their frequent association in colon cancer. As an example, we have here reported the highly significant EIF2 signalling pathway, identified by the contribution of genes EIF2S2 (chr20q11.22), EIF3B (chr7p22.3), EIF3E (Chr8q23.1) and EIF3H (chr8 q24.11). These genes encode eukaryotic initiation factors (eIFs), which are protein complexes involved in the initiation phase of translation. In particular, the EIF2S2 gene encodes one of the subunits of eIF2 (eIF2β). EIF2 is composed of three subunits (α, β, γ), forms a ternary complex with GTP and the initiator tRNA, and binds to a 40S ribosomal

subunit. This eIF3 is formed of 13 subunits (a-m) and participates in the assembly of the 40S ribosomal subunit on mRNA, thus forming a functional pre-initiation complex. It has been reported that eIF3 binds to the 5' untranslated region of a specific set of mRNAs involved in cell growth control (Lee et al., 2015), and that the increased expression of EIF3H gene potentiates colorectal cancer growth and invasiveness (Pittman et al., 2010). Interestingly, another eukaryotic translation initiation factor, eIF6, is included in the list of Over-PositiveT and Fitness-OverT genes located in Chr20. This result is in agreement with our previous data by microarray transcriptomics (Condorelli et al., 2018) and Gandin et al. (Gandin et al., 2008) have already reported that this factor is rate-limiting in translation, growth and transformation. Another significant pathway identified through the analysis of Over-PositiveT and Fitness-OverT is the "Cleavage and Polyadenylation of Pre-mRNA", that contains multiple subunits of CPSF and CSTF complexes. Our analysis identified CPSF1, CPSF4 and CSTF1 among Fitness-OverT genes. Interestingly, Yang et al. have recently reported that a knockdown of CSPF4 in human colorectal cancer cell lines inhibited cell proliferation, migration, invasion and stemness maintenance. Moreover, ectopic overexpression of CPSF4 enhanced tumour growth in mouse models with colon cancer xenografts (Yang et al., 2019). Behan et al. classified CPSF1, CPSF4 and CSTF1 among context-specific fitness genes. In other words, those genes are required for cell fitness in specific molecular or histological contexts, in contrast with "core fitness" genes, that play an essential role in all cell types(Behan et al., 2019).

This observation can be important in the selection of drug targets because the inhibition of context-specific fitness genes has a reduced likelihood of inducing toxic effects in healthy tissues.

In conclusion, our data suggest the hypothesis that an overexpression of specific cancer fitness genes might play a significant role in the functional selection of gained chromosomal arms during cancer evolution. Therefore, Fitness-OverT, as defined in the present paper, might represent candidate targets for a silencing transcriptional therapy in tumours bearing specific BCNGs. These genes have already been shown to be functionally involved in cancer proliferation by the CRISPR-Cas9 inactivation, but the anticancer effects of their partial and combined downregulation remain to be investigated.

5.3 GLIOBASTOMA MULTIFORME (GBM)

Glioblastoma multiforme (GBM) is a malignant primary brain tumour showing a complicated genomic profiling as revealed by the Cancer Genome Atlas project which has involved around 200 human tumour samples (Parsons et al., 2008). Second-generation technologies have focused on recurrent genetic alterations, leading to classify molecularly distinct entities and subtypes (Brennan et al., 2013; Verhaak et al., 2010b)

In this thesis, a wide-transcriptomic and integrated bioinformatic analysis of 33 TCGA cancer types by using cytogenetic data (SNP-arrays) has identified GBM as the first tumours to be characterized by the highest percentage of 7-gain (67% trisomy or tetrasomy of wchr7), followed by KIRP (55.47%), COAD CIN (46.90%), READ CIN (45.09%). These results have oriented the analysis to the deep understanding of the impact that chromosome 7-gain gives to tumorigenesis of GBM. In our analysis involving 592 GBM SNP array samples the gain of chromosome (chr) 7 occurs for 72% in p arm and for 74% in q arm. Another relevant chromosomal aberration is the loss of chr10 (p:74%, q:80%) and the gain of chr19 (p:38%, q:25%). The chromosome 19 loss is usually more frequent in secondary rather than primary GBM (Alifieris et al., 2015; Wilson et al., 2014). Verhaak et al. have seen a co-occurring of chromosome 7 amplification and chromosome 10 loss in 100% of the Classical subtypes (Verhaak et al., 2010b). Exactly because the cytogenetics of GBM is so various and complex, at the beginning we have tested the significant association of chr7-gain and chr19-gain with the well-known amplification of EGFR region (7pFA+)(Brennan et al., 2013). Chr19-gain shows a strong association with 7p-FA con p value < 0.0001 and relative risk of 2.071. Applying the strategy model to investigate broad copy number alteration in GBM cohort we could distinguish 4 groups characterized by the presence of chr7-gain or chr7-disomy in association with EGFR focal amplification (7pFA+ or 7pFA-): 1) Chr7-gain/ 7p-FA+ (n = 40), 2) Chr7-gain / 7p-FA- (n=42), 3) Chr-7disomy/7pFA- (n=15) and 4) Chr7-disomy /7pFA+ (n=8).

The classical subtype is predominant in those groups characterized by 7p-FA+. Verhaak et al 2010 have found EGFR amplification (7p-FA+) in 97% of classical subtype performing a copy number alteration analysis. (Verhaak et al., 2010b)

Systematic transcriptomic analysis revealed an enrichment of transcripts class (the so-called OverT) which are over-expressed in Chr7-gain/ 7p-FA+ group. Precisely, we found 467 OverT genes and, among these, 40 genes were located on chr7p11.2. Regard chr19-gain and 7p-FA also associated with chr7-gain, we found 238 OverT genes. The genes cooperation was

assessed, and main genes were identified. It is relevant the involvement of biological functions in mitochondria based on genes cooperation chr19/chr7/7-FA+. CHCHD2, NIPSNAP2 are genes both located on 7p11.2, while NDUFA13 and NDUFB7 are genes both located on chr19p13. CHCHD2, NIPSNAP2, NDUFA13 and NDUFB7 are over-expressed genes as well as shared genes between chr7-gain /7-FA+ and chr19-gain groups. Among these genes there is a cooperation based on gene dosage-effect which has important repercussions within mitochondria. It has been observed how changes in mitochondrial function or structural changes are a characteristic at different levels in GBM (Arismendi-Morillo, 2011; Katsetos et al., 2013). Malignant glioma cells because of energy deprivation undergo necrosis rather than apoptosis (Steinbach et al., 2003) CHCHD2 is an anti-apoptotic protein localized on the mitochondria and it interacts with Bcl-xL. NDUFA13 is a cell death regulatory gene that promotes apoptosis and a negative regulator of cell growth, and it is also involved in mitochondrial metabolism. These and more observations will be required to shed light on the complex mechanisms bringing to tumorigenesis in GBM. According to literature, the most recurrent genetic alterations are epidermal growth factor receptor (EGFR) up-regulation, phosphate and tensin homologue (PTEN), isocitrate dehydrogenase 1 (IDH1) and p53 mutations (Brennan et al., 2013; Wilson et al., 2014). Analysing 60 GBM samples (WES data matched with RNAseq data, SNParrays) we found that EGFR point mutation characterizes the Chr7-gain FA + Group. There is a mutual exclusion between TP53 points mutations and EGFR mutation in the chr7-gain FA- group.

6. REFERENCES

Acs, G., Lawton, T. J., Rebbeck, T. R., LiVolsi, V. A., & Zhang, P. J. (2001). Differential Expression of E-Cadherin in Lobular and Ductal Neoplasms of the Breast and Its Biologic and Diagnostic Implications. *American Journal of Clinical Pathology*, *115*(1), 85–98. https://doi.org/10.1309/FDHX-L92R-BATQ-2GE0

Ai, H., Barrera, J. E., Meyers, A. D., Shroyer, K. R., & Varella-Garcia, M. (2001). Chromosomal aneuploidy precedes morphological changes and supports multifocality in head and neck lesions. *Laryngoscope*, *111*(10), 1853–1858. https://doi.org/10.1097/00005537-200110000-00034

Akhtar-Zaidi, B., Cowper-Sallari, R., Corradin, O., Saiakhova, A., Bartels, C. F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M. F., Willis, J., Moore, J. H., Tesar, P. J., Laframboise, T., Markowitz, S., Lupien, M., & Scacheri, P. C. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. *Science*, *336*(6082), 736–739. https://doi.org/10.1126/science.1217277

Alsaleem, M., Toss, M. S., Joseph, C., Aleskandarany, M., Kurozumi, S., Alshankyty, I., Ogden, A., Rida, P. C. G., Ellis, I. O., Aneja, R., Green, A. R., Mongan, N. P., & Rakha, E. A. (2019). The molecular mechanisms underlying reduced E-cadherin expression in invasive ductal carcinoma of the breast: high throughput analysis of large cohorts. *Modern Pathology*, *32*(7), 967–976. https://doi.org/10.1038/s41379-019-0209-9

Amiel, A., Gronich, N., Yukla, M., Suliman, S., Josef, G., Gaber, E., Drori, G., Fejgin, M. D., & Lishner, M. (2005). Random aneuploidy in neoplastic and pre-neoplastic diseases, multiple myeloma, and monoclonal gammopathy. *Cancer Genetics and Cytogenetics*, *162*(1), 78–81. https://doi.org/10.1016/j.cancergencyto.2005.03.006

Aqeilan, R. I., Donati, V., Palamarchuk, A., Trapasso, F., Kaou, M., Pekarsky, Y., Sudol, M., & Croce, C. M. (2005). WW domain-containing proteins, WWOX and YAP, compete for interaction with ErbB-4 and modulate its transcriptional function. *Cancer Research*, *65*(15), 6764–6772. https://doi.org/10.1158/0008-5472.CAN-05-1150

Aure, M. R., Steinfeld, I., Baumbusch, L. O., Liestøl, K., Lipson, D., Nyberg, S., Naume, B., Sahlberg, K. K., Kristensen, V. N., Børresen-Dale, A. L., Lingjærde, O. C., & Yakhini, Z. (2013). Identifying In-Trans Process Associated Genes in Breast Cancer by Integrated Analysis of Copy Number and Expression Data. *PLoS ONE*, *8*(1), e53014. https://doi.org/10.1371/journal.pone.0053014

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K. K., Carter, S. L., Frederick, A. M., Lawrence, M. S., Sivachenko, A. Y., Sougnez, C., Zou, L., Cortes, M. L., Fernandez-Lopez, J. C., Peng, S., Ardlie, K. G., Auclair, D., Bautista-Piña, V., Duke, F., Francis, J., Jung, J., … Meyerson, M. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, *486*(7403), 405–409. https://doi.org/10.1038/nature11154

Barbouti, A., Stankiewicz, P., Nusbaum, C., Cuomo, C., Cook, A., Höglund, M., Johansson, B., Hagemeijer, A., Park, S. S., Mitelman, F., Lupski, J. R., & Fioretos, T. (2004). The Breakpoint Region of the Most Common Isochromosome, i(17q), in Human Neoplasia Is Characterized by a Complex Genomic Architecture with Large, Palindromic, Low-Copy

Repeats. *American Journal of Human Genetics*, *74*(1), 1–10. https://doi.org/10.1086/380648

Barresi, V., Castorina, S., Musso, N., Capizzi, C., Luca, T., Privitera, G., & Condorelli, D. F. (2017). Chromosomal instability analysis and regional tumor heterogeneity in colon cancer. *Cancer Genetics*, *210*, 9–21. https://doi.org/10.1016/j.cancergen.2016.11.001

Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., Ansari, R., Harper, S., Jackson, D. A., McRae, R., Pooley, R., Wilkinson, P., van der Meer, D., Dow, D., Buser-Doepner, C., … Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, *568*(7753), 511–516. https://doi.org/10.1038/s41586-019-1103-9

Ben-David, U., & Amon, A. (2020). Context is everything: aneuploidy in cancer. In *Nature Reviews Genetics* (Vol. 21, Issue 1, pp. 44–62). Nature Research. https://doi.org/10.1038/s41576-019-0171-x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J. C., Huang, J. H., Alexander, S., Du, J., Kau, T., Thomas, R. K., Shah, K., Soto, H., Perner, S., Prensner, J., Debiasi, R. M., Demichelis, F., … Sellers, W. R. (2007). Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20007–20012. https://doi.org/10.1073/pnas.0710052104

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., McHenry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y. J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., … Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, *463*(7283), 899–905. https://doi.org/10.1038/nature08822

Berx, G., Cleton-Jansen, A. M., Strumane, K., De Leeuw, W. J. F., Nollet, F., Van Roy, F., & Cornelisse, C. (1996). E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene*, *13*(9), 1919–1925.

Borcherding, N., Cole, K., Kluz, P., Jorgensen, M., Kolb, R., Bellizzi, A., & Zhang, W. (2018). Re-Evaluating E-Cadherin and β-Catenin: A Pan-Cancer Proteomic Approach with an Emphasis on Breast Cancer. *American Journal of Pathology*, *188*(8), 1910–1920. https://doi.org/10.1016/j.ajpath.2018.05.003

Bouteille, N., Driouch, K., Hage, P. E., Sin, S., Formstecher, E., Camonis, J., Lidereau, R., & Lallemand, F. (2009). Inhibition of the Wnt/B-catenin pathway by the WWOX tumor suppressor protein. *Oncogene*, *28*(28), 2569–2580. https://doi.org/10.1038/onc.2009.120

Boveri, T. (1902). Uber mehrpolige Mitosen als Mittes zur Analyse des Zellkerns. *Verhandlungen Der Physikalisch-Medizinischen Gesellschaft Zu Wurzburg*, *35*, 67–90. http://10e.devbio.com/article.php?id=24

Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.

R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhim, R., Bernard, B., Wu, C. J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S. A., … McLendon, R. (2013). The somatic genomic landscape of glioblastoma. *Cell*, *155*(2), 462. https://doi.org/10.1016/j.cell.2013.09.034

Brenner, A. J., & Aldaz, C. M. (1995). Chromosome 9p Allelic Loss and p16/CDKN2 in Breast Cancer and Evidence of p16 Inactivation in Immortal Breast Epithelial Cells. *Cancer Research*, *55*(13).

Bürger, H., de Boer, M., van Diest, P. J., & Korsching, E. (2013). Chromosome 16q loss- a genetic key to the understanding of breast carcinogenesis. In *Histology and Histopathology* (Vol. 28, Issue 3, pp. 311–320). https://doi.org/10.14670/HH-28.311

Cardoso, J., Molenaar, L., De Menezes, R. X., Van Leerdam, M., Rosenberg, C., Möslein, G., Sampson, J., Morreau, H., Boer, J. M., & Fodde, R. (2006). Chromosomal instability in MYH- and APC-mutant adenomatous polyps. *Cancer Research*, *66*(5), 2514–2519. https://doi.org/10.1158/0008-5472.CAN-05-2407

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

Chen, J., Rajasekaran, M., Xia, H., Kong, S. N., Deivasigamani, A., Sekar, K., Gao, H., Swa, H. L., Gunaratne, J., Ooi, L. L., Xie, T., Hong, W., & Hui, K. M. (2018). CDK 1-mediated BCL 9 phosphorylation inhibits clathrin to promote mitotic Wnt signalling. *The EMBO Journal*, *37*(20), e99395. https://doi.org/10.15252/embj.201899395

Chin, K., DeVries, S., Fridlyand, J., Spellman, P. T., Roydasgupta, R., Kuo, W. L., Lapuk, A., Neve, R. M., Qian, Z., Ryder, T., Chen, F., Feiler, H., Tokuyasu, T., Kingsley, C., Dairkee, S., Meng, Z., Chew, K., Pinkel, D., Jain, A., … Gray, J. W. (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*, *10*(6), 529–541. https://doi.org/10.1016/j.ccr.2006.10.009

Choi, Y. J., Pinto, M. M., Hao, L., & Riba, A. K. (2008). Interobserver variability and aberrant E-cadherin immunostaining of lobular neoplasia and infiltrating lobular carcinoma. *Modern Pathology*, *21*(10), 1224–1237. https://doi.org/10.1038/modpathol.2008.106

Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., Bowlby, R., Shen, H., Hayat, S., Fieldhouse, R., Lester, S. C., Tse, G. M. K., Factor, R. E., Collins, L. C., Allison, K. H., … Perou, C. M. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, *163*(2), 506–519. https://doi.org/10.1016/j.cell.2015.09.033

Citri, A., Skaria, K. B., & Yarden, Y. (2003). The deaf and the dumb: The biology of ErbB-2 and ErbB-3. In *Experimental Cell Research* (Vol. 284, Issue 1, pp. 54–65). Academic Press Inc. https://doi.org/10.1016/S0014-4827(02)00101-5

Clevers, H., Loh, K. M., & Nusse, R. (2014). An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. In *Science* (Vol. 346, Issue 6205). American Association for the Advancement of Science. https://doi.org/10.1126/science.1248012

Cohen, A. J., Saiakhova, A., Corradin, O., Luppino, J. M., Lovrenert, K., Bartels, C. F., Morrow, J. J., Mack, S. C., Dhillon, G., Beard, L., Myeroff, L., Kalady, M. F., Willis, J., Bradner, J. E., Keri, R. A., Berger, N. A., Pruett-Miller, S. M., Markowitz, S. D., & Scacheri, P. C. (2017). Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14400

Condorelli, D. F., Spampinato, G., Valenti, G., Musso, N., Castorina, S., & Barresi, V. (2018). Positive Caricature Transcriptomic Effects Associated with Broad Genomic Aberrations in Colorectal Cancer. *Scientific Reports*, *8*(1), 14826. https://doi.org/10.1038/s41598-018-32884-3

Condorelli, D., Privitera, A. P., & Barresi, V. (2019). Chromosomal density of cancer up-regulated genes, aberrant enhancer activity and cancer fitness genes are associated with transcriptional cis-effects of broad copy number gains in colorectal cancer. *International Journal of Molecular Sciences*, *20*(18), 4652. https://doi.org/10.3390/ijms20184652

Cowell, J. K., LaDuca, J., Rossi, M. R., Burkhardt, T., Nowak, N. J., & Matsui, S. I. (2005). Molecular characterization of the t(3;9) associated with immortalization in the MCF10A cell line. *Cancer Genetics and Cytogenetics*, *163*(1), 23–29. https://doi.org/10.1016/j.cancergencyto.2005.04.019

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Aparicio, S., Brenton, J. D., Ellis, I., … Caldas, C. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, *486*(7403), 346–352. https://doi.org/10.1038/nature10983

Dabbs, D. J., Schnitt, S. J., Geyer, F. C., Weigelt, B., Baehner, F. L., Decker, T., Eusebi, V., Fox, S. B., Ichihara, S., Lakhani, S. R., Palacios, J., Rakha, E., Richardson, A. L., Schmitt, F. C., Tan, P. H., Tse, G. M., Vincent-Salomon, A., Ellis, I. O., Badve, S., & Reis-Filho, J. S. (2013). Lobular neoplasia of the breast revisited with emphasis on the role of e-cadherin immunohistochemistry. *American Journal of Surgical Pathology*, *37*(7), e1–e11. https://doi.org/10.1097/PAS.0b013e3182918a2b

Davoli, T., Uno, H., Wooten, E. C., & Elledge, S. J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*, *355*(6322), eaaf8399. https://doi.org/10.1126/science.aaf8399

Davoli, T., Xu, A. W., Mengwasser, K. E., Sack, L. M., Yoon, J. C., Park, P. J., & Elledge, S. J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*, *155*(4), 948. https://doi.org/10.1016/j.cell.2013.10.011

De Leeuw, W. J. F., Berx, G., Vos, C. B. J., Peterse, J. L., Van De Vijver, M. J., Litvinov, S., Van Roy, F., Cornelisse, C. J., & Cleton-Jansen, A. M. (1997). Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *Journal of Pathology*, *183*(4), 404–411. https://doi.org/10.1002/(SICI)1096-9896(199712)183:4<404::AID-PATH1148>3.0.CO;2-9

Debnath, J., Muthuswamy, S. K., & Brugge, J. S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. In *Methods* (Vol. 30, Issue 3, pp. 256–268). Academic Press Inc. https://doi.org/10.1016/S1046-2023(03)00032-X

Desmedt, C., Zoppoli, G., Gundem, G., Pruneri, G., Larsimont, D., Fornili, M., Fumagalli, D., Brown, D., Rothé, F., Vincent, D., Kheddoumi, N., Rouas, G., Majjaj, S., Brohée, S., Van Loo, P., Maisonneuve, P., Salgado, R., Van Brussel, T., Lambrechts, D., … Sotiriou, C. (2016). Genomic Characterization of Primary Invasive Lobular Breast Cancer. *Journal of Clinical Oncology*, *34*(16), 1872–1880. https://doi.org/10.1200/JCO.2015.64.0334

Doak, S. H., Jenkins, G. J. S., Parry, E. M., Griffiths, A. P., Baxter, J. N., & Parry, J. M. (2004). Differential expression of the MAD2, BUB1 and HSP27 genes in Barrett's oesophagus - Their association with aneuploidy and neoplastic progression. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *547*(1–2), 133–144. https://doi.org/10.1016/j.mrfmmm.2003.12.009

Doebele, R. C., Drilon, A., Paz-Ares, L., Siena, S., Shaw, A. T., Farago, A. F., Blakely, C. M., Seto, T., Cho, B. C., Tosi, D., Besse, B., Chawla, S. P., Bazhenova, L., Krauss, J. C., Chae, Y. K., Barve, M., Garrido-Laguna, I., Liu, S. V., Conkling, P., … Demetri, G. D. (2020). Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials. *The Lancet Oncology*, *21*(2), 271–282. https://doi.org/10.1016/S1470-2045(19)30691-6

Dooley, T. P., Mattern, V. L., Moore, C. M., Porter, P. A., Robinson, E. S., & VandeBerg, J. L. (1993). Cell lines derived from μLtraviolet radiation-induced benign melanocytic nevi in Monodelphis domestica exhibit cytogenetic aneuploidy. *Cancer Genetics and Cytogenetics*, *71*(1), 55–66. https://doi.org/10.1016/0165-4608(93)90202-W

Dowsett, M., Sestak, I., Lopez-Knowles, E., Sidhu, K., Dunbier, A. K., Cowens, J. W., Ferree, S., Storhoff, J., Schaper, C., & Cuzick, J. (2013). Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *Journal of Clinical Oncology*, *31*(22), 2783–2790. https://doi.org/10.1200/JCO.2012.46.1558

Droufakou, S., Deshmane, V., Roylance, R., Hanby, A., Tomlinson, I., & Hart, I. R. (2001). Multiple ways of silencing E-cadherin gene expression in lobular carcinoma of the breast. *International Journal of Cancer*, *92*(3), 404–408. https://doi.org/10.1002/ijc.1208

Duensing, S., & Münger, K. (2004). Mechanisms of genomic instability in human cancer: Insights from studies with human papillomavirus oncoproteins. *International Journal of Cancer*, *109*(2), 157–162. https://doi.org/10.1002/ijc.11691

Dutrillaux, B., Gerbault-Seureau, M., & Zafrani, B. (1990). Characterization of chromosomal anomalies in human breast cancer. A comparison of 30 paradiploid cases with few chromosome changes. *Cancer Genetics and Cytogenetics*, *49*(2), 203–217. https://doi.org/10.1016/0165-4608(90)90143-X

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 14863–14868. https://doi.org/10.1073/pnas.95.25.14863

Elsarraj, H. S., Hong, Y., Valdez, K. E., Michaels, W., Hook, M., Smith, W. P., Chien, J., Herschkowitz, J. I., Troester, M. A., Beck, M., Inciardi, M., Gatewood, J., May, L., Cusick, T., McGinness, M., Ricci, L., Fan, F., Tawfik, O., Marks, J. R., … Behbod, F. (2015). Expression profiling of in vivo ductal carcinoma in situ progression models

identified B cell lymphoma-9 as a molecular driver of breast cancer invasion. *Breast Cancer Research*, *17*(1), 128. https://doi.org/10.1186/s13058-015-0630-z

Euhus, D. M., Timmons, C. F., & Tomlinson, G. E. (2002). ETV6-NTRK3 - Trk-ing the primary event in human secretory breast cancer. In *Cancer Cell* (Vol. 2, Issue 5, pp. 347–348). Cell Press. https://doi.org/10.1016/S1535-6108(02)00184-8

Farabegoli, F., Hermsen, M. A. J. A., Ceccarelli, C., Santini, D., Weiss, M. M., Meijer, G. A., & Van Diest, P. J. (2004). Simultaneous chromosome 1q gain and 16q loss is associated with steroid receptor presence and low proliferation in breast carcinoma. *Modern Pathology*, *17*(4), 449–455. https://doi.org/10.1038/modpathol.3800059

Fehrmann, R. S. N., Karjalainen, J. M., Krajewska, M., Westra, H. J., Maloney, D., Simeonov, A., Pers, T. H., Hirschhorn, J. N., Jansen, R. C., Schultes, E. A., Van Haagen, H. H. H. B. M., De Vries, E. G. E., Te Meerman, G. J., Wijmenga, C., Van Vugt, M. A. T. M., & Franke, L. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nature Genetics*, *47*(2), 115–125. https://doi.org/10.1038/ng.3173

Ferguson, D. O., & Alt, F. W. (2001). DNA double strand break repair and chromosomal translocation: Lessons from animal models. In *Oncogene* (Vol. 20, Issue 40 REV. ISS. 4, pp. 5572–5579). Oncogene. https://doi.org/10.1038/sj.onc.1204767

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., & Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, *136*(5), E359–E386. https://doi.org/10.1002/ijc.29210

Filipović, A., Gronau, J. H., Green, A. R., Wang, J., Vallath, S., Shao, D., Rasul, S., Ellis, I. O., Yagüe, E., Sturge, J., & Coombes, R. C. (2011). Biological and clinical implications of nicastrin expression in invasive breast cancer. *Breast Cancer Research and Treatment*, *125*(1), 43–53. https://doi.org/10.1007/s10549-010-0823-1

Gamallo, C., Palacios, J., Benito, N., Limeres, M. A., Pizarro, A., Suárez, A., Pastrana, F., Cano, A., & Calero, F. (1996). Expression of E-cadherin in 230 infiltrating ductal breast carcinoma: Relationship to clinicopathological features. *International Journal of Oncology*, *9*(6), 1207–1212. https://doi.org/10.3892/ijo.9.6.1207

Gandin, V., Miluzio, A., Barbieri, A. M., Beugnet, A., Kiyokawa, H., Marchisio, P. C., & Biffo, S. (2008). Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. *Nature*, *455*(7213), 684–688. https://doi.org/10.1038/nature07267

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, *6*(269), pl1. https://doi.org/10.1126/scisignal.2004088

Geisenberger, C., Mock, A., Warta, R., Rapp, C., Schwager, C., Korshunov, A., Nied, A. K., Capper, D., Brors, B., Jungk, C., Jones, D., Collins, V. P., Ichimura, K., Bäcklund, L. M., Schnabel, E., Mittelbron, M., Lahrmann, B., Zheng, S., Verhaak, R. G. W., … Herold-Mende, C. (2015). Molecular profiling of long-term survivors identifies a subgroup of glioblastoma characterized by chromosome 19/20 co-gain. *Acta Neuropathologica*, *130*(3), 419–434. https://doi.org/10.1007/s00401-015-1427-y

Gnant, M., Filipits, M., Greil, R., Stoeger, H., Rudas, M., Bago-Horvath, Z., Mlineritsch, B.,

Kwasny, W., Knauer, M., Singer, C., Jakesz, R., Dubsky, P., Fitzal, F., Bartsch, R., Steger, G., Balic, M., Ressler, S., Cowens, J. W., Storhoff, J., … Nielsen, T. O. (2014). Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: Using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Annals of Oncology*, *25*(2), 339–345. https://doi.org/10.1093/annonc/mdt494

Grabenstetter, A., Mohanty, A. S., Rana, S., Zehir, A., Brannon, A. R., D'Alfonso, T. M., DeLair, D. F., Tan, L. K., & Ross, D. S. (2020). E-cadherin immunohistochemical expression in invasive lobular carcinoma of the breast: correlation with morphology and CDH1 somatic alterations. *Human Pathology*, *102*, 44–53. https://doi.org/10.1016/j.humpath.2020.06.002

Grady, W. M. (2004). Genomic instability and colon cancer. In *Cancer and Metastasis Reviews* (Vol. 23, Issues 1–2, pp. 11–27). Springer. https://doi.org/10.1023/A:1025861527711

Graus-Porta, D., Beerli, R. R., Daly, J. M., & Hynes, N. E. (1997). ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO Journal*, *16*(7), 1647–1655. https://doi.org/10.1093/emboj/16.7.1647

Hamilton, S., Bosman, F. T., & Boffetta, P. (2010). *Carcinoma of the Colon and Rectum. In Bosman, F.T.C.F., Hruban, R.H. and Theise, N.D., Eds., World Health Organization Classification of Tumours of the Digestive System, International Agency for Research on Cancer, Lyon, 134-146. - References - Sci*.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. In *Cell* (Vol. 100, Issue 1, pp. 57–70). Elsevier. https://doi.org/10.1016/S0092-8674(00)81683-9

Hallmarks of cancer: The next generation, 144 Cell 646 (2011). https://doi.org/10.1016/J.CELL.2011.02.013

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F. P., Rissland, O. S., Durocher, D., Angers, S., & Moffat, J. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, *163*(6), 1515–1526. https://doi.org/10.1016/j.cell.2015.11.015

Hawthorn, L., Luce, J., Stein, L., & Rothschild, J. (2010). Integration of transcript expression, copy number and LOH analysis of infiltrating ductal carcinoma of the breast. *BMC Cancer*, *10*, 460. https://doi.org/10.1186/1471-2407-10-460

Heberle, H., Meirelles, V. G., da Silva, F. R., Telles, G. P., & Minghim, R. (2015). InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, *16*(1). https://doi.org/10.1186/s12859-015-0611-3

Hu, C., Diévart, A., Lupien, M., Calvo, E., Tremblay, G., & Jolicoeur, P. (2006). Overexpression of activated murine Notch1 and Notch3 in transgenic mice blocks mammary gland development and induces mammary tumors. *American Journal of Pathology*, *168*(3), 973–990. https://doi.org/10.2353/ajpath.2006.050416

Huiping, C., Sigurgeirsdottir, J. R., Jonasson, J. G., Eiriksdottir, G., Johannsdottir, J. T., Egilsson, V., & Ingvarsson, S. (1999). Chromosome alterations and E-cadherin gene mutations in human lobular breast cancer. *British Journal of Cancer*, *81*(7), 1103–1110.

https://doi.org/10.1038/sj.bjc.6690815

Hungermann, D., Schmidt, H., Natrajan, R., Tidow, N., Poos, K., Reis-Filho, J. S., Brandt, B., Buerger, H., & Korsching, E. (2011). Influence of whole arm loss of chromosome 16q on gene expression patterns in oestrogen receptor-positive, invasive breast cancer. *Journal of Pathology*, *224*(4), 517–528. https://doi.org/10.1002/path.2938

Jenkins, R. B., Blair, H., Ballman, K. V., Giannini, C., Arusell, R. M., Law, M., Flynn, H., Passe, S., Felten, S., Brown, P. D., Shaw, E. G., & Buckner, J. C. (2006). A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma. *Cancer Research*, *66*(20), 9852–9861. https://doi.org/10.1158/0008-5472.CAN-06-1796

Jeschke, U., Mylonas, I., Kuhn, C., Shabani, N., Kunert-Keil, C., Schindlbeck, C., Gerber, B., & Friese, K. (2007). Expression of E-cadherin in human ductal breast cancer carcinoma in situ, invasive carcinomas, their lymph node metastases, their distant metastases, carcinomas with recurrence and in recurrence. *Anticancer Research*, *27*(4 A), 1969–1974.

Jiang, M., Kang, Y., Sewastianik, T., Wang, J., Tanton, H., Alder, K., Dennis, P., Xin, Y., Wang, Z., Liu, R., Zhang, M., Huang, Y., Loda, M., Srivastava, A., Chen, R., Liu, M., & Carrasco, R. D. (2020). BCL9 provides multi-cellular communication properties in colorectal cancer by interacting with paraspeckle proteins. *Nature Communications*, *11*(1), 19. https://doi.org/10.1038/s41467-019-13842-7

Jung, J. G., Stoeck, A., Guan, B., Wu, R. C., Zhu, H., Blackshaw, S., Shih, I. M., & Wang, T. L. (2014). Notch3 Interactome Analysis Identified WWP2 as a Negative Regulator of Notch3 Signaling in Ovarian Cancer. *PLoS Genetics*, *10*(10), e1004751. https://doi.org/10.1371/journal.pgen.1004751

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database*, *2011*. https://doi.org/10.1093/database/bar030

Knezevich, S. R., McFadden, D. E., Tao, W., Lim, J. F., & Sorensen, P. H. B. (1998). A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. *Nature Genetics*, *18*(2), 184–187. https://doi.org/10.1038/ng0298-184

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R., Wilson, R. K., Ally, A., Balasundaram, M., Butterfield, Y. S. N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H. J. E., … Palchik, J. D. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61–70. https://doi.org/10.1038/nature11412

Kokalj-Vokac, N., Alemeida, A., Gerbault-Seureau, M., Malfoy, B., & Dutrillaux, B. (1993). Two-color FISH characterization of i(1q) and der(1;16) in human breast cancer cells. *Genes, Chromosomes and Cancer*, *7*(1), 8–14. https://doi.org/10.1002/gcc.2870070103

Kolde, R. (2012). Pheatmap: pretty heatmaps, R package v. 16. *(R Foundation for Statistical Computing)*.

Laé, M., Fréneaux, P., Sastre-Garau, X., Chouchane, O., Sigal-Zafrani, B., & Vincent-Salomon, A. (2009). Secretory breast carcinomas with ETV6-NTRK3 fusion gene

belong to the basal-like carcinoma spectrum. *Modern Pathology*, *22*(2), 291–298. https://doi.org/10.1038/modpathol.2008.184

Lang, S. J., Schmiech, M., Hafner, S., Paetz, C., Werner, K., Gaafary, M. El, Schmidt, C. Q., Syrovets, T., & Simmet, T. (2020). Chrysosplenol d, a flavonol from artemisia annua, induces erk1/2-mediated apoptosis in triple negative human breast cancer cells. *International Journal of Molecular Sciences*, *21*(11), 1–19. https://doi.org/10.3390/ijms21114090

Larson, D. E., Harris, C. C., Chen, K., Koboldt, D. C., Abbott, T. E., Dooling, D. J., Ley, T. J., Mardis, E. R., Wilson, R. K., & Ding, L. (2012). Somaticsniper: Identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*, *28*(3), 311–317. https://doi.org/10.1093/bioinformatics/btr665

Lee, A. S. Y., Kranzusch, P. J., & Cate, J. H. D. (2015). EIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*, *522*(7554), 111–114. https://doi.org/10.1038/nature14267

Lin, S. Y., Xia, W., Wang, J. C., Kwong, K. Y., Spohn, B., Wen, Y., Pestell, R. G., & Hung, M. C. (2000). β-catenin, a novel prognostic marker for breast cancer: Its roles in cyclin D1 expression and cancer progression. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4262–4266. https://doi.org/10.1073/pnas.060025397

Ling, H., Sylvestre, J. R., & Jolicoeur, P. (2013). Cyclin D1-dependent induction of luminal inflammatory breast tumors by activated Notch3. *Cancer Research*, *73*(19), 5963–5973. https://doi.org/10.1158/0008-5472.CAN-13-0409

Linggi, B., & Carpenter, G. (2006). ErbB receptors: new insights on mechanisms and biology. In *Trends in Cell Biology* (Vol. 16, Issue 12, pp. 649–656). Trends Cell Biol. https://doi.org/10.1016/j.tcb.2006.10.008

Liu, Q., Schwaller, J., Kutok, J., Cain, D., Aster, J. C., Williams, I. R., & Gilliland, D. G. (2000). Signal transduction and transforming properties of the TEL-TRKC fusions associated with t(12;15)(p13;q25) in congenital fibrosarcoma and acute myelogenous leukemia. *EMBO Journal*, *19*(8), 1827–1838. https://doi.org/10.1093/emboj/19.8.1827

Liu, Y., Sethi, N. S., Hinoue, T., Schneider, B. G., Cherniack, A. D., Sanchez-Vega, F., Seoane, J. A., Farshidfar, F., Bowlby, R., Islam, M., Kim, J., Chatila, W., Akbani, R., Kanchi, R. S., Rabkin, C. S., Willis, J. E., Wang, K. K., McCall, S. J., Mishra, L., … Laird, P. W. (2018). Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell*, *33*(4), 721-735.e8. https://doi.org/10.1016/j.ccell.2018.03.010

Malik, N., Yan, H., Moshkovich, N., Palangat, M., Yang, H., Sanchez, V., Cai, Z., Peat, T. J., Jiang, S., Liu, C., Lee, M., Mock, B. A., Yuspa, S. H., Larson, D., Wakefield, L. M., & Huang, J. (2019). The transcription factor CBFB suppresses breast cancer through orchestrating translation and transcription. *Nature Communications*, *10*(1), 2071. https://doi.org/10.1038/s41467-019-10102-6

Mandel, C. R., Bai, Y., & Tong, L. (2008). Protein factors in pre-mRNA 3′-end processing. In *Cellular and Molecular Life Sciences* (Vol. 65, Issues 7–8, pp. 1099–1122). https://doi.org/10.1007/s00018-007-7474-3

Marambaud, P., Shioi, J., Serban, G., Georgakopoulos, A., Sarner, S., Nagy, V., Baki, L., Wen, P., Efthimiopoulos, S., Shao, Z., Wisniewski, T., & Robakis, N. K. (2002). A presenilin-1/γ-secretase cleavage releases the E-cadherin intracellular domain and regulates disassembly of adherens junctions. *EMBO Journal*, *21*(8), 1948–1956. https://doi.org/10.1093/emboj/21.8.1948

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., & Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Research*, *28*(11), 1747–1756. https://doi.org/10.1101/gr.239244.118

McCart Reed AE, Kutasovic JR, Lakhani SR, S. P. (2015). Invasive lobular carcinoma of the breast: Morphology, biomarkers and 'omics. In *Breast Cancer Research* (Vol. 17, Issue 1, p. 12). https://doi.org/10.1186/s13058-015-0519-x

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

Medina, D. (2002). Biological and molecular characteristics of the premalignant mouse mammary gland. In *Biochimica et Biophysica Acta - Reviews on Cancer* (Vol. 1603, Issue 1, pp. 1–9). Biochim Biophys Acta. https://doi.org/10.1016/S0304-419X(02)00053-7

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., & Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology*, *12*(4). https://doi.org/10.1186/gb-2011-12-4-r41

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*, *49*(D1), D394–D403. https://doi.org/10.1093/nar/gkaa1106

Michaut, M., Chin, S. F., Majewski, I., Severson, T. M., Bismeijer, T., De Koning, L., Peeters, J. K., Schouten, P. C., Rueda, O. M., Bosma, A. J., Tarrant, F., Fan, Y., He, B., Xue, Z., Mittempergher, L., Kluin, R. J. C., Heijmans, J., Snel, M., Pereira, B., … Bernards, R. (2016). Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Scientific Reports*, *6*, 18517. https://doi.org/10.1038/srep18517

Mieszczanek, J., van Tienen, L. M., Ibrahim, A. E. K., Winton, D. J., & Bienz, M. (2019). Bcl9 and Pygo synergise downstream of Apc to effect intestinal neoplasia in FAP mouse models. *Nature Communications*, *10*(1), 724. https://doi.org/10.1038/s41467-018-08164-z

Mitelman, F., Johansson, B., & Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. In *Nature Reviews Cancer* (Vol. 7, Issue 4, pp. 233–245). Nat Rev Cancer. https://doi.org/10.1038/nrc2091

Moasser, M. M. (2007). The oncogene HER2: Its signaling and transforming functions and its role in human cancer pathogenesis. In *Oncogene* (Vol. 26, Issue 45, pp. 6469–6487). Oncogene. https://doi.org/10.1038/sj.onc.1210477

Mollen, E. W. J., Ient, J., Tjan-Heijnen, V. C. G., Boersma, L. J., Miele, L., Smidt, M. L., &

Vooijs, M. A. G. G. (2018). Moving breast cancer therapy up a notch. In *Frontiers in Oncology* (Vol. 8, p. 518). https://doi.org/10.3389/fonc.2018.00518

Mueller, C., Haymond, A., Davis, J. B., Williams, A., & Espina, V. (2018). Protein biomarkers for subtyping breast cancer and implications for future research. *Expert Review of Proteomics*, *15*(2), 131–152. https://doi.org/10.1080/14789450.2018.1421071

Mukherjee, N., Bhattacharya, N., Alam, N., Roy, A., Roychoudhury, S., & Panda, C. K. (2012). Subtype-specific alterations of the Wnt signaling pathway in breast cancer: Clinical and prognostic significance. *Cancer Science*, *103*(2), 210–220. https://doi.org/10.1111/j.1349-7006.2011.02131.x

Mukherjee, N., Dasgupta, H., Bhattacharya, R., Pal, D., Roy, R., Islam, S., Alam, N., Biswas, J., Roy, A., Roychoudhury, S., & Panda, C. K. (2016). Frequent inactivation of MCC/CTNNBIP1 and overexpression of phospho-beta-cateninY654 are associated with breast carcinoma: Clinical and prognostic significance. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1862*(9), 1472–1484. https://doi.org/10.1016/j.bbadis.2016.05.009

Mukherjee, N., & Panda, C. K. (2020). Wnt/β-Catenin Signaling Pathway as Chemotherapeutic Target in Breast Cancer: An Update on Pros and Cons. In *Clinical Breast Cancer* (Vol. 20, Issue 5, pp. 361–370). https://doi.org/10.1016/j.clbc.2020.04.004

Muthuswami, M., Ramesh, V., Banerjee, S., Viveka Thangaraj, S., Periasamy, J., Bhaskar Rao, D., Barnabas, G. D., Raghavan, S., & Ganesan, K. (2013). Breast Tumors with Elevated Expression of 1q Candidate Genes Confer Poor Clinical Outcome and Sensitivity to Ras/PI3K Inhibition. *PLoS ONE*, *8*(10), e77553. https://doi.org/10.1371/journal.pone.0077553

Nascimento, R. G. do, & Otoni, K. M. (2020). Histological and molecular classification of breast cancer: what do we know? *Mastology*, *30*, 1–8. https://doi.org/10.29289/25945394202020200024

Network, C. G. A. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *New England Journal of Medicine*, *372*(26), 2481–2498. https://doi.org/10.1056/nejmoa1402121

Pandis, N., Bardi, G., Jin, Y., Dietrich, C., Johansson, B., Andersen, J., Mandahl, N., Mitelman, F., & Heim, S. (1994). Unbalanced t(1;16) as the sole karyotypic abnormality in a breast carcinoma and its lymph node metastasis. In *Cancer Genetics and Cytogenetics* (Vol. 75, Issue 2, pp. 158–159). https://doi.org/10.1016/0165-4608(94)90172-4

Pandis, Nikos, Heim, S., Bardi, G., Idvall, I., Mandahl, N., & Mitelman, F. (1992). Whole-arm t(1;16) and i(1q) as sole anomalies identify gain of 1 q as a primary chromosomal abnormality in breast cancer. *Genes, Chromosomes and Cancer*, *5*(3), 235–238. https://doi.org/10.1002/gcc.2870050310

Pandis, Nikos, Jin, Y., Gorunova, L., Petersson, C., Bardi, G., Idvall, I., Johansson, B., Ingvar, C., Mandahl, N., Mitelman, F., & Heim, S. (1995). Chromosome analysis of 97 primary breast carcinomas: Identification of eight karyotypic subgroups. *Genes, Chromosomes and Cancer*, *12*(3), 173–185. https://doi.org/10.1002/gcc.2870120304

Pareja, F., Weigelt, B., & Reis-Filho, J. S. (2021). Problematic breast tumors reassessed in light of novel molecular data. *Modern Pathology*, *34*(i), 38–47. https://doi.org/10.1038/s41379-020-00693-7

Peltonen, H. M., Haapasalo, A., Hiltunen, M., Kataja, V., Kosma, V. M., & Mannermaa, A. (2013). γ-Secretase Components As Predictors of Breast Cancer Outcome. *PLoS ONE*, *8*(11), e79249. https://doi.org/10.1371/journal.pone.0079249

Pereira, B., Chin, S. F., Rueda, O. M., Vollan, H. K. M., Provenzano, E., Bardwell, H. A., Pugh, M., Jones, L., Russell, R., Sammut, S. J., Tsui, D. W. Y., Liu, B., Dawson, S. J., Abraham, J., Northen, H., Peden, J. F., Mukherjee, A., Turashvili, G., Green, A. R., … Caldas, C. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11479

Perou, C. M., Sørile, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Ress, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, Ø., Pergammenschlkov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., & Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), 747–752. https://doi.org/10.1038/35021093

Picornell, A. C., Echavarria, I., Alvarez, E., López-Tarruella, S., Jerez, Y., Hoadley, K., Parker, J. S., Del Monte-Millán, M., Ramos-Medina, R., Gayarre, J., Ocaña, I., Cebollero, M., Massarrah, T., Moreno, F., García Saenz, J. A., Gómez Moreno, H., Ballesteros, A., Ruiz Borrego, M., Perou, C. M., & Martin, M. (2019). Breast cancer PAM50 signature: Correlation and concordance between RNA-Seq and digital multiplexed gene expression technologies in a triple negative breast cancer series. *BMC Genomics*, *20*(1), 1–11. https://doi.org/10.1186/s12864-019-5849-0

Pittman, A. M., Naranjo, S., Jalava, S. E., Twiss, P., Ma, Y., Olver, B., Lloyd, A., Vijayakrishnan, J., Qureshi, M., Broderick, P., van Wezel, T., Morreau, H., Tuupanen, S., Aaltonen, L. A., Alonso, M. E., Manzanares, M., Gavilán, A., Visakorpi, T., Gómez-Skarmeta, J. L., & Houlston, R. S. (2010). Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a Cis-acting regulator of EIF3H. *PLoS Genetics*, *6*(9). https://doi.org/10.1371/journal.pgen.1001126

Pollack, J. R., Sørlie, T., Perou, C. M., Rees, C. A., Jeffrey, S. S., Lonning, P. E., Tibshirani, R., Botstein, D., Børresen-Dale, A. L., & Brown, P. O. (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12963–12968. https://doi.org/10.1073/pnas.162471999

Pospiech, K., Pluciennik, E., & Bednarek, A. K. (2018). WWOX tumor suppressor gene in breast cancer, a historical perspective and future directions. In *Frontiers in Oncology* (Vol. 8, Issue AUG, p. 345). https://doi.org/10.3389/fonc.2018.00345

Pradeep, C. R., Köstler, W. J., Lauriola, M., Granit, R. Z., Zhang, F., Jacob-Hirsch, J., Rechavi, G., Nair, H. B., Hennessy, B. T., Gonzalez-Angulo, A. M., Tekmal, R. R., Ben-Porath, I., Mills, G. B., Domany, E., & Yarden, Y. (2012). Modeling ductal carcinoma in situ: A HER2-Notch3 collaboration enables luminal filling. *Oncogene*, *31*(7), 907–917. https://doi.org/10.1038/onc.2011.279

Privitera, A. P., Barresi, V., & Condorelli, D. F. (2021). Aberrations of chromosomes 1 and

16 in breast cancer: A framework for cooperation of transcriptionally dysregulated genes. *Cancers*, *13*(7). https://doi.org/10.3390/cancers13071585

Ptashkin, R. N., Pagan, C., Yaeger, R., Middha, S., Shia, J., O'Rourke, K. P., Berger, M. F., Wang, L., Cimera, R., Wang, J., Klimstra, D. S., Saltz, L., Ladanyi, M., Zehir, A., & Hechtman, J. F. (2017). Chromosome 20q amplification defines a subtype of microsatellite stable, left-sided colon cancers with wild-type RAS/RAF and better overall survival. *Molecular Cancer Research*, *15*(6), 708–713. https://doi.org/10.1158/1541-7786.MCR-16-0352

Ran, Y., Hossain, F., Pannuti, A., Lessard, C. B., Ladd, G. Z., Jung, J. I., Minter, L. M., Osborne, B. A., Miele, L., & Golde, T. E. (2017). γ-Secretase inhibitors in cancer clinical trials are pharmacologically and functionally distinct. *EMBO Molecular Medicine*, *9*(7), 950–966. https://doi.org/10.15252/emmm.201607265

Ried, T., Heselmeyer-Haddad, K., Blegen, H., Schr&#xfffd;ck, E., & Auer, G. (1999). Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: A phenotype/genotype correlation. *Genes, Chromosomes and Cancer*, *25*(3), 195–204. https://doi.org/10.1002/(SICI)1098-2264(199907)25:3<195::AID-GCC1>3.0.CO;2-8

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Rose, J. T., Moskovitz, E., Boyd, J. R., Gordon, J. A., Bouffard, N. A., Fritz, A. J., Illendula, A., Bushweller, J. H., Lian, J. B., Stein, J. L., Zaidi, S. K., & Stein, G. S. (2020). Inhibition of the RUNX1-CBFβ transcription factor complex compromises mammary epithelial cell identity: A phenotype potentially stabilized by mitotic gene bookmarking. *Oncotarget*, *11*(26), 2512–2530. https://doi.org/10.18632/oncotarget.27637

Roy, D. M., Walsh, L. A., Desrichard, A., Huse, J. T., Wu, W., Gao, J. J., Bose, P., Lee, W., & Chan, T. A. (2016). Integrated Genomics for Pinpointing Survival Loci within Arm-Level Somatic Copy Number Alterations. *Cancer Cell*, *29*(5), 737–750. https://doi.org/10.1016/j.ccell.2016.03.025

Roylance, R., Gorman, P., Papior, T., Wan, Y. L., Ives, M., Watson, J. E., Collins, C., Wortham, N., Langford, C., Fiegler, H., Carter, N., Gillett, C., Sasieni, P., Pinder, S., Hanby, A., & Tomlinson, I. (2006). A comprehensive study of chromosome 16q in invasive ductal and lobular breast carcinoma using array CGH. *Oncogene*, *25*(49), 6544–6553. https://doi.org/10.1038/sj.onc.1209659

Rueda, O. M., Sammut, S. J., Seoane, J. A., Chin, S. F., Caswell-Jin, J. L., Callari, M., Batra, R., Pereira, B., Bruna, A., Ali, H. R., Provenzano, E., Liu, B., Parisien, M., Gillett, C., McKinney, S., Green, A. R., Murphy, L., Purushotham, A., Ellis, I. O., … Curtis, C. (2019). Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature*, *567*(7748), 399–404. https://doi.org/10.1038/s41586-019-1007-8

Russnes, H. G., Lingjærde, O. C., Børresen-Dale, A. L., & Caldas, C. (2017). Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. In *American*

Journal of Pathology. https://doi.org/10.1016/j.ajpath.2017.04.022

Russnes, H. G., Vollan, H. K. M., Lingjærde, O. C., Krasnitz, A., Lundin, P., Naume, B., Sørlie, T., Borgen, E., Rye, I. H., Langerød, A., Chin, S. F., Teschendorff, A. E., Stephens, P. J., Månér, S., Schlichting, E., Baumbusch, L. O., Kåresen, R., Stratton, M. P., Wigler, M., … Børresen-Dale, A. L. (2010). Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Science Translational Medicine*, *2*(38), 38ra47. https://doi.org/10.1126/scitranslmed.3000611

Rye, I. H., Lundin, P., Månér, S., Fjelldal, R., Naume, B., Wigler, M., Hicks, J., Børresen-Dale, A. L., Zetterberg, A., & Russnes, H. G. (2015). Quantitative multigene FISH on breast carcinomas identifies der(1;16)(q10;p10) as an early event in luminal A tumors. *Genes Chromosomes and Cancer*, *54*(4), 235–248. https://doi.org/10.1002/gcc.22237

Sack, L. M., Davoli, T., Li, M. Z., Li, Y., Xu, Q., Naxerova, K., Wooten, E. C., Bernardi, R. J., Martin, T. D., Chen, T., Leng, Y., Liang, A. C., Scorsone, K. A., Westbrook, T. F., Wong, K. K., & Elledge, S. J. (2018). Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell*, *173*(2), 499-514.e23. https://doi.org/10.1016/j.cell.2018.02.037

Scaltriti, M., & Baselga, J. (2006). The epidermal growth factor receptor pathway: A model for targeted therapy. In *Clinical Cancer Research* (Vol. 12, Issue 18, pp. 5268–5272). Clin Cancer Res. https://doi.org/10.1158/1078-0432.CCR-05-1554

Schuchardt, B. J., Bhat, V., Mikles, D. C., McDonald, C. B., Sudol, M., & Farooq, A. (2013). Molecular origin of the binding of WWOX tumor suppressor to erbb4 receptor tyrosine kinase. *Biochemistry*, *52*(51), 9223–9236. https://doi.org/10.1021/bi400987k

Shackleford, M. T., Rao, D. M., Bordeaux, E. K., Hicks, H. M., Towers, C. G., Sottnik, J. L., Oesterreich, S., & Sikora, M. J. (2020). Estrogen regulation of mTOR signaling and mitochondrial function in invasive lobular carcinoma cell lines requires WNT4. *Cancers*, *12*(10), 1–22. https://doi.org/10.3390/cancers12102931

Shi, Y., Di Giammartino, D. C., Taylor, D., Sarkeshik, A., Rice, W. J., Yates, J. R., Frank, J., & Manley, J. L. (2009). Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. *Molecular Cell*, *33*(3), 365–376. https://doi.org/10.1016/j.molcel.2008.12.028

Solimini, N. L., Xu, Q., Mermel, C. H., Liang, A. C., Schlabach, M. R., Luo, J., Burrows, A. E., Anselmo, A. N., Bredemeyer, A. L., Li, M. Z., Beroukhim, R., Meyerson, M., & Elledge, S. J. (2012). Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science*, *336*(6090), 104–109. https://doi.org/10.1126/science.1219580

Soneson, C. (2014). Compcoder-An R package for benchmarking differential expression methods for RNA-seq data. *Bioinformatics*, *30*(17), 2517–2518. https://doi.org/10.1093/bioinformatics/btu324

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E., & Børresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(19), 10869–10874. https://doi.org/10.1073/pnas.191367098

Spandidos, A., Wang, X., Wang, H., Dragnev, S., Thurber, T., & Seed, B. (2008). A comprehensive collection of experimentally validated primers for Polymerase Chain Reaction quantitation of murine transcript abundance. *BMC Genomics*, *9*(1), 633. https://doi.org/10.1186/1471-2164-9-633

Spandidos, A., Wang, X., Wang, H., & Seed, B. (2009). PrimerBank: A resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Research*, *38*(SUPPL.1), D792–D799. https://doi.org/10.1093/nar/gkp1005

Srihari, S., Kalimutho, M., Lal, S., Singla, J., Patel, D., Simpson, P. T., Khanna, K. K., & Ragan, M. A. (2016). Understanding the functional impact of copy number alterations in breast cancer using a network modeling approach. *Molecular BioSystems*, *12*(3), 963–972. https://doi.org/10.1039/c5mb00655d

Subik, K., Lee, J. F., Baxter, L., Strzepek, T., Costello, D., Crowley, P., Xing, L., Hung, M. C., Bonfiglio, T., Hicks, D. G., & Tang, P. (2010). The expression patterns of ER, PR, HER2, CK5/6, EGFR, KI-67 and AR by immunohistochemical analysis in breast cancer cell lines. *Breast Cancer: Basic and Clinical Research*, *4*(1), 35–41. https://doi.org/10.1177/117822341000400004

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., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Tauriello, D. V. F., Haegebarth, A., Kuper, I., Edelmann, M. J., Henraat, M., Canninga-van Dijk, M. R., Kessler, B. M., Clevers, H., & Maurice, M. M. (2010). Loss of the Tumor Suppressor CYLD Enhances Wnt/β-Catenin Signaling through K63-Linked Ubiquitination of Dvl. *Molecular Cell*, *37*(5), 607–619. https://doi.org/10.1016/j.molcel.2010.01.035

Taylor, A. M., Shih, J., Ha, G., Gao, G. F., Zhang, X., Berger, A. C., Schumacher, S. E., Wang, C., Hu, H., Liu, J., Lazar, A. J., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., Hutter, C. M., Sofia, H. J., Tarnuzzer, R., … Meyerson, M. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*, *33*(4), 676-689.e3. https://doi.org/10.1016/j.ccell.2018.03.007

Teixeira, M. R., Pandis, N., & Heim, S. (2002). Cytogenetic clues to breast carcinogenesis. In *Genes Chromosomes and Cancer* (Vol. 33, Issue 1, pp. 1–16). https://doi.org/10.1002/gcc.1206

Teixeira, M. R., Pandis, N., & Heim, S. (2015). Tumors of the breast. In L. John Wiley & Sons (Ed.), *Cancer Cytogenetics: Fourth Edition* (pp. 426–446). https://doi.org/10.1002/9781118795569.ch16

Tentler, J. J., Lang, J., Capasso, A., Kim, D. J., Benaim, E., Lee, Y. B., Eisen, A., Bagby, S. M., Hartman, S. J., Yacob, B. W., Gittleman, B., Pitts, T. M., Pelanda, R., Eckhardt, S. G., & Diamond, J. R. (2020). RX-5902, a novel β-catenin modulator, potentiates the efficacy of immune checkpoint inhibitors in preclinical models of triple-negative breast Cancer. *BMC Cancer*, *20*(1), 1063. https://doi.org/10.1186/s12885-020-07500-1

Tian, B., & Manley, J. L. (2016). Alternative polyadenylation of mRNA precursors. In *Nature Reviews Molecular Cell Biology* (Vol. 18, Issue 1, pp. 18–30). Nature Publishing Group.

https://doi.org/10.1038/nrm.2016.116

Tognon, C., Garnett, M., Kenward, E., Kay, R., Morrison, K., & Sorensen, P. H. B. (2001). The Chimeric Protein Tyrosine Kinase ETV6-NTRK3 Requires both Ras-Erk1/2 and PI3-Kinase-Akt Signaling for Fibroblast Transformation. *Cancer Research*, *61*(24).

Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. In *Wspolczesna Onkologia* (Vol. 1A, pp. A68–A77). https://doi.org/10.5114/wo.2014.47136

Tsarouha, H., Pandis, N., Bardi, G., Teixeira, M. R., Andersen, J. A., & Heim, S. (1999). Karyotypic E evolution in breast carcinomas with i(1)(q10) and der(1;16)(q10;p10) as the primary chromosome abnormality. *Cancer Genetics and Cytogenetics*, *113*(2), 156–161. https://doi.org/10.1016/S0165-4608(99)00016-3

Upender, M. B., Habermann, J. K., McShane, L. M., Korn, E. L., Barrett, J. C., Difilippantonio, M. J., & Ried, T. (2004). Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer Research*, *64*(19), 6941–6949. https://doi.org/10.1158/0008-5472.CAN-04-0474

Van Der Meer, D., Barthorpe, S., Yang, W., Lightfoot, H., Hall, C., Gilbert, J., Francies, H. E., & Garnett, M. J. (2019). Cell Model Passports - a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Research*, *47*(D1), D923–D929. https://doi.org/10.1093/nar/gky872

van Tienen, L. M., Mieszczanek, J., Fiedler, M., Rutherford, T. J., & Bienz, M. (2017). Constitutive scaffolding of multiple Wnt enhanceosome components by legless/BCL9. *ELife*, *6*. https://doi.org/10.7554/eLife.20882

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., Mesirov, J. P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B. A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., … Hayes, D. N. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, *17*(1), 98–110. https://doi.org/10.1016/j.ccr.2009.12.020

Villa, J. C., Chiu, D., Brandes, A. H., Escorcia, F. E., Villa, C. H., Maguire, W. F., Hu, C. J., deStanchina, E., Simon, M. C., Sisodia, S. S., Scheinberg, D. A., & Li, Y. M. (2014). Nontranscriptional role of hif-1α in activation of γ-secretase and notch signaling in breast cancer. *Cell Reports*, *8*(4), 1077–1092. https://doi.org/10.1016/j.celrep.2014.07.028

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Smits, A. M. M., & Bos, J. L. (1988). Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine*, *319*(9), 525–532. https://doi.org/10.1056/nejm198809013190901

Wai, D. H., Knezevich, S. R., Lucas, T., Jansen, B., Kay, R. J., & Sorensen, P. H. B. (2000). The ETV6-NTRK3 gene fusion encodes a chimeric protein tyrosine kinase that transforms NIH3T3 cells. *Oncogene*, *19*(7), 906–915. https://doi.org/10.1038/sj.onc.1203396

Walter, W., Sánchez-Cabo, F., & Ricote, M. (2015). GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics (Oxford, England)*,

31(17), 2912–2914. https://doi.org/10.1093/bioinformatics/btv300

Wang, M., Zhao, Y., & Zhang, B. (2015). Efficient Test and Visualization of Multi-Set Intersections. *Nature Publishing Group*. https://doi.org/10.1038/srep16923

Wang, X., & Seed, B. (2003). A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Research*, *31*(24). https://doi.org/10.1093/nar/gng154

Weaver, A., & Cleveland, D. W. (2006). Does aneuploidy cause cancer? In *Current Opinion in Cell Biology* (Vol. 18, Issue 6, pp. 658–667). Elsevier Ltd. https://doi.org/10.1016/j.ceb.2006.10.002

Weigelt, B., Horlings, H. M., Kreike, B., Hayes, M. M., Hauptmann, M., Wessels, L. F. A., De Jong, D., Van De Vijver, M. J., Van't Veer, L. J., & Peterse, J. L. (2008). Refinement of breast cancer classification by molecular characterization of histological special types. *Journal of Pathology*, *216*(2), 141–150. https://doi.org/10.1002/path.2407

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Sander, C., Stuart, J. M., Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chu, A., … Kling, T. (2013). The cancer genome atlas pan-cancer analysis project. In *Nature Genetics* (Vol. 45, Issue 10, pp. 1113–1120). https://doi.org/10.1038/ng.2764

Wieduwilt, M. J., & Moasser, M. M. (2008). The epidermal growth factor receptor family: Biology driving targeted therapeutics. In *Cellular and Molecular Life Sciences* (Vol. 65, Issue 10, pp. 1566–1584). Cell Mol Life Sci. https://doi.org/10.1007/s00018-008-7440-8

Xu, X., Zhang, M., Xu, F., & Jiang, S. (2020). Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. In *Molecular Cancer* (Vol. 19, Issue 1, p. 165). https://doi.org/10.1186/s12943-020-01276-5

Xu, Y., Liao, S., Wang, L., Wang, Y., Wei, W., Su, K., Tu, Y., & Zhu, S. (2020). Galeterone sensitizes breast cancer to chemotherapy via targeting MNK/eIF4E and β-catenin. *Cancer Chemotherapy and Pharmacology*, *Nov 7*. https://doi.org/10.1007/s00280-020-04195-w

Yang, Q., Fan, W., Zheng, Z., Lin, S., Liu, C., Wang, R., Li, W., Zuo, Y., Sun, Y., Hu, S., Chen, M., Guo, P., Pan, J., Tian, C., Zhu, T., Diao, C., Hao, J., Yu, W., Li, L., … Wu, X. (2019). Cleavage and polyadenylation specific factor 4 promotes colon cancer progression by transcriptionally activating hTERT. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1866*(10), 1533–1543. https://doi.org/10.1016/j.bbamcr.2019.07.001

Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., Lawrence, M. S., Zhang, C. Z., Wala, J., Mermel, C. H., Sougnez, C., Gabriel, S. B., Hernandez, B., Shen, H., Laird, P. W., Getz, G., Meyerson, M., & Beroukhim, R. (2013). Pan-cancer patterns of somatic copy number alteration. *Nature Genetics*, *45*(10), 1134–1140. https://doi.org/10.1038/ng.2760

Zheng, J. (2013). Oncogenic chromosomal translocations and human cancer (Review). In *Oncology Reports* (Vol. 30, Issue 5, pp. 2011–2019). Spandidos Publications. https://doi.org/10.3892/or.2013.2677

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C.,

& Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Zientek-Targosz, H., Kunnev, D., Hawthorn, L., Venkov, M., Matsui, S. I., Cheney, R. T., & Ionov, Y. (2008). Transformation of MCF-10A cells by random mutagenesis with frameshift mutagen ICR191: A model for identifying candidate breast-tumor suppressors. *Molecular Cancer*, *7*. https://doi.org/10.1186/1476-4598-7-51

7. APPENDIX

Publications

- Privitera AP, Barresi V, Condorelli DF. **Aberrations of Chromosomes 1 and 16 in Breast Cancer: A Framework for Cooperation of Transcriptionally Dysregulated Genes**. Cancers (Basel). 2021 Mar 30;13(7):1585. doi: 10.3390/cancers13071585. PMID: 33808143; PMCID: PMC8037453.

- Barresi V, Di Bella V, Andriano N, Privitera AP, Bonaccorso P, La Rosa M, Iachelli V, Spampinato G, Pulvirenti G, Scuderi C, Condorelli DF, Lo Nigro L. **NUP-98 Rearrangements Led to the Identification of Candidate Biomarkers for Primary Induction Failure in Pediatric Acute Myeloid Leukemia.** Int J Mol Sci. 2021 Apr 27;22(9):4575. doi: 10.3390/ijms22094575. PMID: 33925480; PMCID: PMC8123909.

- Condorelli DF, Privitera AP, Barresi V. ***Chromosomal Density of Cancer Up-Regulated Genes, Aberrant Enhancer Activity and Cancer Fitness Genes Are* Associated with Transcriptional Cis-Effects of Broad Copy Number Gains in Colorectal Cancer.** Int J Mol Sci. 2019 Sep 19;20(18). pii: E4652. doi: 10.3390/ijms20184652.

Conferences and Posters

- 32nd Annual Conference of Italian Association of Cell Cultures (AICC) **From Single Gene Analysis to Single Cell Profiling: A new Era for Genomic Medicine** October 1st – 2nd, 2019 – Catanzaro, Italy.
  POSTER TITLE: ABERRANT ENHANCER LOCI AND CANCER FITNESS PROPERTIES ARE ASSOCIATED WITH ANEUPLOIDY-DEPENDENT TRANSCRIPTIONAL EFFECTS.
  Anna P. Privitera, Giorgia Spampinato, Vincenza Barresi, Daniele F. Condorelli

- 61° SIB MEETING Virtual Edition 23-24 September 2021 Scientific Program
  3 Posters
  1) Identification of cancer driver genes affected by broad copy number aberrations of chromosome 1 and 16 in breast cancer
  Privitera A.P, Scuderi C., Di Bella V., Barresi V. and Condorelli D.F.

  2) Genes associated to NUP98-rearrangements are involved in primary chemoresistance of pediatric Acute Myeloid Leukemia
  Virginia Di Bella, Luca Lo Nigro, Nellina Andriano, Anna Provvidenza Privitera, Paola Bonaccorso, Manuela La Rosa, Valeria Iachelli, Giorgia Spampinato, Giulio Pulvirenti, Chiara Scuderi, Daniele F. Condorelli and Vincenza Barresi.

  3) **Aneuploidy and Translation Initiation Factors in Colorectal Cancer**

Chiara Scuderi, Anna P. Privitera, Virginia Di Bella, Vincenza Barresi, Daniele F. Condorelli

- 53° CONGRESSO NAZIONALE SIBIOC - Medicina di Laboratorio- La Medicina di laboratorio nella medicina di prossimità telemedicina, ruoli e competenze (13 Ottobre 2021).
  1) Identification of cancer driver genes affected by broad copy number aberrations of chromosome 1 and 16 in breast cancer
  Privitera A.P, Scuderi C., Di Bella V., Barresi V. and Condorelli D.F.

- 1st Workshop of the SIB group "Tumor Biochemistry" From genes to metabolites through proteins: dealing with human health and disease
  25th February 2022, Università Milano Bicocca
  3 Posters
  1) γ-secretase complex in breast cancer cells bearing aberrations of chromosome 1 and 16.
  Privitera Anna P., Scuderi Chiara. , Di Bella Virginia, Barresi Vincenza. and Condorelli Daniele.F.

  2) **Transcriptional dysregulation of EIF3H and EIF2S2 in aneuploid colon cancer**
  Chiara Scuderi, Anna P. Privitera, Virginia Di Bella, Morgana D'Amico, Luca Lanzanò, Vincenza Barresi, Daniele F. Condorelli.

  3) **Serine Protease Inhibitor Kazal-type 2 in normal hematopoietic stem cells and leukemia**
  Virginia Di Bella, Luca Lo Nigro, Anna Provvidenza Privitera, Paola Bonaccorso, Chiara Scuderi, Daniele F. Condorelli and Vincenza Barresi.