

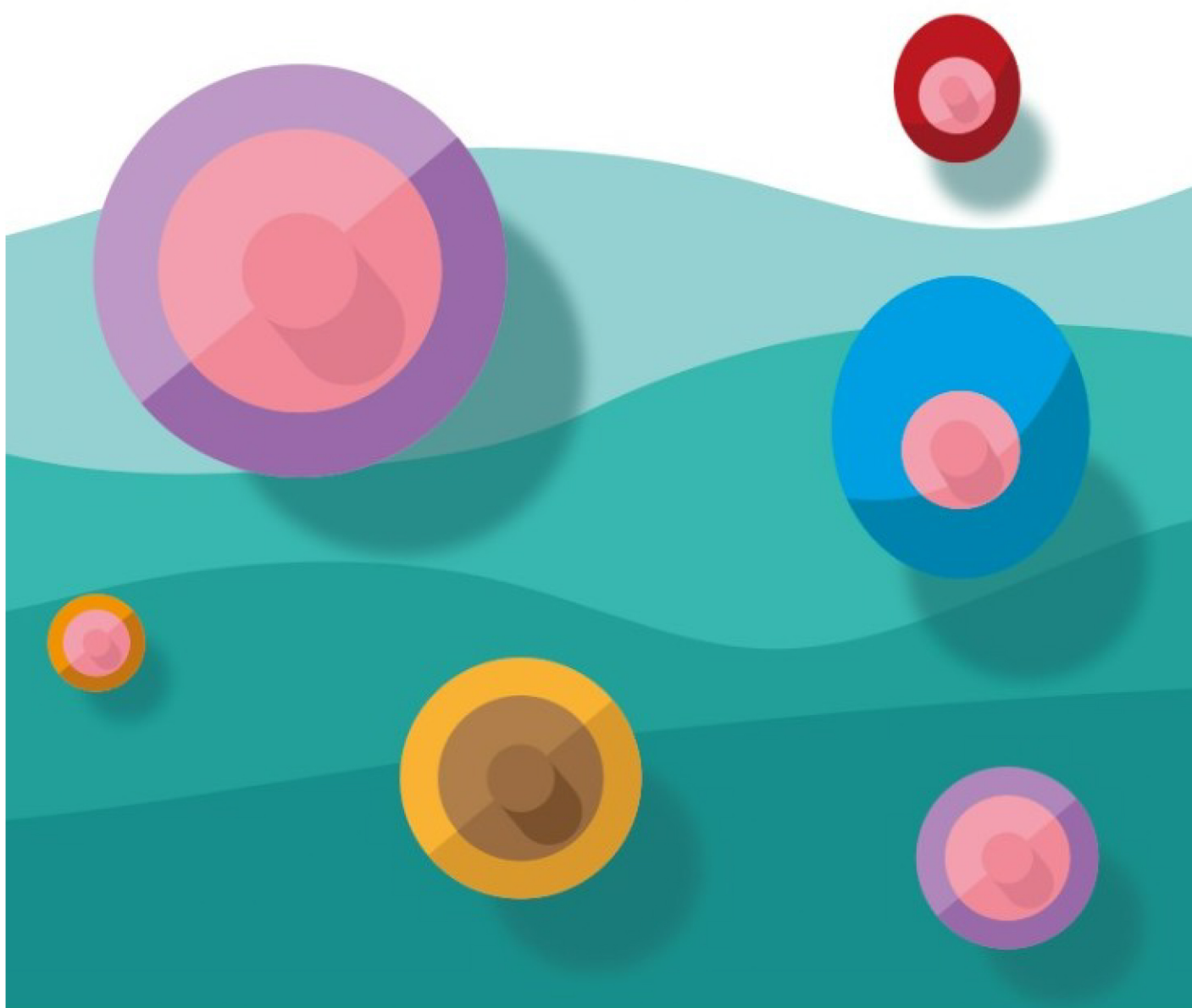
TESIS DOCTORAL



PAISAJE MOLECULAR DEL TRATAMIENTO NEOADYUVANTE EN EL CÁNCER DE MAMA:

PREDICCIÓN DE LA RESPUESTA, RESISTENCIA Y RECAÍDA A DISTANCIA

María del Rosario Chica Parrado



UNIVERSIDAD
DE MÁLAGA




Directores:
Emilio Alba Conejo
Isabel Barragán Mallofret

Programa de Doctorado en Biología Molecular y Celular
Universidad de Málaga
Facultad de Ciencias
Málaga, 2021



UNIVERSIDAD
DE MÁLAGA

AUTOR: María Rosario Chica Parrado

 <https://orcid.org/0000-0001-5007-7063>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es





TESIS DOCTORAL

**PAISAJE MOLECULAR DEL TRATAMIENTO
NEOADYUVANTE EN CÁNCER DE MAMA:
PREDICCIÓN DE LA RESPUESTA,
RESISTENCIA Y RECAÍDA A DISTANCIA**

MARÍA DEL ROSARIO CHICA PARRADO

Directores: Emilio Alba Conejo y M^a Isabel Barragán Mallofret
Tutor: Francisco Cánovas

Programa de Doctorado en Biología Celular y Molecular
Facultad de Ciencias. Universidad de Málaga

Málaga, España 2021





DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña MARÍA DEL ROSARIO CHICA PARRADO

Estudiante del programa de doctorado BIOLÓGIA MOLECULAR Y CELULAR de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: PAISAJE MOLECULAR DEL TRATAMIENTO NEOADYUVANTE EN CÁNCER DE MAMA: PREDICCIÓN DE LA RESPUESTA, RESISTENCIA Y RECAÍDA A DISTANCIA.

Realizada bajo la tutorización de FRANCISCO M CÁNOVAS y dirección de EMLIO ALBA CONEJO E ISABEL BARRAGÁN MALLOFRET (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 25 de MAYO de 2021



Por la presente, **D. Emilio Alba Conejo**, Catedrático de Medicina de la Universidad de Málaga, director del Centro de investigaciones Médico-Sanitarias (CIMES) de la Universidad de Málaga, **D^a. M^a Isabel Barragán Mallofret**, Investigadora Principal, Programa Nicolás Monardes (Acción de Vinculación de investigadores a las Unidades de Gestión Clínica (UGC) del Servicio Andaluz de Salud), en la UGC Oncología Intercentros, profesora externa en el Departamento de Fisiología y Farmacología, Instituto Karolinska y **D. Francisco M Cánovas**, Catedrático de Bioquímica y Biología Molecular de la Universidad de Málaga

CERTIFICAN que **D^a. María del Rosario Chica Parrado** ha obtenido y estudiado personalmente bajo nuestra dirección y supervisión los datos necesarios para la realización de su Tesis Doctoral, titulada: “Paisaje molecular del tratamiento neoadyuvante en cáncer de mama: predicción de la respuesta, resistencia y recaída a distancia”, que consideramos tiene el contenido y rigor científico necesario para ser sometido al superior juicio de la Comisión que nombre la Universidad de Málaga para optar a grado de Doctor.

Y para que así conste a los efectos oportunos, firman el presente documento en 26 de mayo de 2021.

*Para mi tata,
por enseñarme a quererme*

*Para mi hermana,
por enseñarme a desaprender*

*Para mis padres,
por el apoyo incondicional*

Agradecimientos

Todo el mundo piensa que lo importante de una tesis son los resultados obtenidos o lo bien planteada que esté la discusión. Bien sabe el que lo ha vivido desde dentro que una tesis es el compendio de muchas cosas, de saber escuchar los consejos (muchas gracias **Paco Cánovas**), de agradecer que te den la oportunidad de estar (eternamente agradecida, **Emilio Alba**), de ser y desarrollarte (**Martina Álvarez** y **Rocío Lavado**, gracias), de aprender a no rendirme (**Isabel Barragan**, eres toda una cheerleader), de fomentar el pensamiento crítico (**Emilio** e **Isabel**, muy orgullosa de haber sido vuestra alumna), a **Juan Luis Onieva**, mil gracias por tu inestimable ayuda, y por último, tengo que agradecerme a mí misma por permitirme disfrutar del camino. Para esto último tengo/necesito/debo dar las gracias a mucha gente, que por suerte (para mí), han estado de manera inquebrantable. Agradezco a todo el **personal del LBMC** por aguantarme durante este tiempo, sé que no es tarea fácil. En especial quiero agradecer a **Angy**, lo mejor que me llevo, mi hermana, aunque nuestros genes no digan lo mismo...Eres inspiración ♥, a **Alix**, gracias por aguantar todas mis tonterías como una auténtica campeona, a **Elee**, gracias por ser tan madre, eres como estar en casa, guashimodo nos espera, **Aurora ailovyu** and you know it!, a **Souka**, por tanta nobleza, besitos y corazoncitos, a **Alfon**, la mesa echará de menos nuestros golpes de risa (y yo tus ocurrencias), a **Celia** por ser una mezcla de rigurosidad y cachondeo a partes iguales, a **Iñaki**, en poco tiempo me has conquistado, te voy a echar de menos (tutiechocolate?) y a mi **Gemi** (Mari) por alegrar cada uno de los días que he estado en este laboratorio (transferensiah Marii?).

Pero para que una construcción se lleve a cabo de manera sólida es necesario una buena base. Y puedo presumir de cimientos. **Pepis**, una vida sin tus quejas y apoyo incondicional ya no la quiero ♥, **Marino (Marrano)**, eres el hogar a donde sé que siempre puedo acudir y **Juanillo**, el perreo sin ti no es lo mismo (adoro mi portada y al artista que la ha diseñado), os quiero chicos, ¿nos vemos en Tarifa? **Kat**, eres ejemplo para seguir, única e inigualable, loveU. **Alba** y **Laura**, la superación y la fuerza hecha personas, os prometo que os dedico el nobel, palabrita. Y a las 3 de las 4 fantásticas, **Luni**, **Alix** y **Marinetta**, ¡os quiero! Pero sobre todo gracias a mi **padre**, a mi **madre**, a mi **hermana**, a mis **abuelos** y a toda mi familia, sin ellos no estaría donde estoy ni sería lo que soy.

Summary

INTRODUCTION:

Breast cancer is not a single disease, but a group of diseases that, despite being located in the same organ, present great variability in all its components: histopathological characteristics, response to treatment, and survival (1). This variability is observed both within a single tumor (intratumoral heterogeneity) and between tumors from different patients (intertumoral heterogeneity) at different levels: morphological, molecular and genomic (2). From the point of view of intertumoral heterogeneity, the identification of subgroups of patients with homogeneous characteristics would help us to improve our understanding of this disease and, therefore, the patient care.

NEOADJUVANT TREATMENT

Neoadjuvant treatment (NAC) (based on the administration of Anthracyclines +/- Taxanes, and Trastuzumab in those cases with HER2 overexpression) before surgery is widely used in patients with breast cancer both in early stages and in the scenario of the locally advanced disease. The response to NAC, which can be dichotomized into 1) pathological complete response (pCR) when the tumor responds to treatment or 2) presence of residual disease (RD) when the response to treatment is not complete, is crucial for understanding the evolution of the patient after treatment. Achieving pCR after surgery has been shown to be associated with greater overall and disease-free survival. However, the presence of tumor cells or RD is associated with a worse long-term clinical outcome. The response to treatment with NAC is different in the various subtypes of breast cancer, with patients with the subtypes with the most activated proliferation pathways, such as Triple Negative or HER2 +, being the most responsive (3) (4).

Based on the above, the **hypotheses** of this study are that there are certain molecular profiles that differentially predict the response to treatment with NAC in primary breast

cancer in the different subtypes of BC without necessarily being surrogate markers for these. The changes in the molecular profile observed in the primary tumor with respect to RD after treatment with NAC allow the identification of potentially exploitable molecular alterations from the therapeutic point of view. On the other hand, the molecular profile of the residual tumor consistently predicts the molecular profile of the metastatic tumor.

Therefore, the **objectives** of this work are: 1) To determine a predictive signature of pCR that identifies before treatment those patients who will not respond to Neoadjuvant chemotherapy with greater specificity and independent of the subtype than current prediction tools. 2) Identify the common and specific molecular alterations of the primary tumor and the residual tumor after treatment with NAC to understand tumor evolution and to identify specific resistance markers. And 3) Establish the spectrum of molecular variants associated with the first metastatic niche to establish specific metastasis markers and those predictors of it from the primary tumor and the RD.

CHAPTER 1: Prediction of Response to Neoadjuvant Treatment

The overall percentage of patients who obtain pCR after treatment with NAC is around 31% (5); It is, therefore, of vital importance to find markers that identify those patients who are going to obtain pCR from those who do not before treatment.

In this scenario, both clinicopathological variables have been identified (mammographic tumor size, tumor grade, subtypes determined by immunohistochemistry (IHC), armpit lymph node involvement or % of KI67) as well as molecular markers (molecular subtypes defined by PAM50, Integrative Cluster (IntClust), or the TRAR model (Trastuzumab Risk model) that are related to obtaining pCR (1). Regarding clinicopathological variables, tumor size is one of the variables associated with pCR in a distributed cohort in 10.3% HER2-enriched, 18.2% Luminal B, 30.6% Luminal A and 32.7% Basal-Like ($p < 0.001$). Those tumors with a size of cT1-T2 achieve twice the response to NAC than larger tumors (cT3-T4) (2). Another one is tumor histological grade; in a cohort of 560 primary tumors, Lips et al. (2013) demonstrated that a high tumor histological grade (G3)

predicted pCR in ER + / HER2 breast tumors and in triple negative tumors (TN) treated with NAC plus platinum salts (6). These same results were obtained by Diaz-Redondo and colleagues (2019) in a cohort of 259 patients with HER2 + breast tumors (7). The use of the Ki67 variable before treatment as a predictor of pCR after treatment with NAC has been shown to be useful especially in ER- / HER2 + and TN breast tumors, since they are the ones with the highest expression of proliferation-related pathways. In fact, several studies indicate that a high percentage of Ki67 (> 50%) is an important predictor of response to treatment with NAC (8) (9) (10) (11). Determining the histological subtypes based on mainly three markers such as the estrogen receptor (ER), the progesterone receptor (PR) and the overexpression of HER2, also determines subgroups of patients with a specific long-term result after treatment. Those patients with triple negative breast tumors (TN) or with HER2 + will respond better to treatment with NAC, while the subtypes that are positive for hormone receptors (ER and PR) will present a lower percentage of pCR (12) (13) (14) (3) (4) (5).

In the scenario of response predictors based on the study of the molecular biology of tumors, PAM50, as a subtype classifier based on the expression of 50 genes, is postulated to date as the best predictor of response to treatment with NAC given the extensive number of publications that support its predictive capacity. Studies derived from the clinical trials NOAH, CherLOB, CALGB 40,601, PAMELA and BERENICE, among others, conclude that tumors classified as HER2-enriched have a high rate of pCR. PAM50 is also capable of differentially predicting pCR within HER2 + tumors determined by IHC, so that, of them, those that PAM50 classifies as Luminal A subtype, on the contrary, have the lowest pCR ratios (1) (15) (16) (17) (18) (19). When all breast cancer subtypes are included, the Basal-Like and HER2-enriched subtypes are associated with the highest ratio of pCR. Consistent with the aforementioned, the Luminal A subtype (associated in turn to profiles with the least expressed proliferation mechanisms) is the one that obtains the lowest response rate and is therefore the subtype of breast cancer defined by PAM50 that least benefits from NAC treatment (1) (2) (20) (15). On the other hand, the IntClust classification of breast tumors, based on the alteration in the number of copies that determines different subtypes and different clinical outcomes, has been

shown to be useful in predicting the response to NAC in a cohort of 100 patients from the GEICAM / 2006–03 and GEICAM / 2006–14 clinical trials (21). Finally, in HER2 + tumors, a 41–gene signature, called TRAR, demonstrated pCR predictive ability in a study derived from the NeoALTTO clinical trial. In this case, it is possible to obtain specificity and sensitivity information, such that TRAR obtained in its ROC curve an AUC value of 0.73 [22]. Although some of these markers are very promising, such as the molecular subtypes of breast cancer defined by PAM50, there is currently no clear candidate that is used at the clinical level to distinguish responding from non–responding patients.

The objective of the first chapter of this thesis is **to characterize patients at the molecular level before being treated with NAC and to relate the identified differences with the response to treatment** retrospectively. By doing this, and applying our knowledge in the field of Machine Learning and integrating the clinicopathological characteristics, our second objective is **to define a predictive signature of pCR** that improves the predictive value of the current markers.

Study population and methodology

The most relevant results of this study have been carried out with 84 patients with breast cancer, treated with Neoadjuvant treatment. The analyzes have been carried out using the diagnostic biopsy before the treatment and the response to the treatment has been defined in the tissue from the surgery with the Miller & Payne method.

To determine the predictive signatures of pCR, we have defined several cohorts based on provenance. The cohort where the different predictive models have been defined has been called **Cohort_Discovery** and the cohort where we have validated these predictive models has been called **Cohort_Validation**. Both cohorts have been subdivided according to the different breast cancer subtype markers: **Allsubtypes**, which includes all subtypes without distinction of markers, **HER2–**, which includes all patients without HER2 amplification or overexpression (without taking into account the ER receptor), and an **ER +** cohort, which includes ER + / HER2– patients. Therefore, we are going to have 3 sub–

cohorts for each cohort. Similarly, the cohort where the predictive signatures have been defined has been divided into: Cohort_D_Allsubtypes, Cohort_D_HER2- and Cohort_D_ER +, and the cohort where these signatures have been validated has been divided into: Cohort_V_Allsubtypes, Cohort_V_HERV_ER2- and Cohort_V_HERV_ER2-.

In all patients, the main known markers of response to treatment with NAC were determined, such as tumor size, histological tumor grade, Ki67 percentage, immunohistochemical subtypes of breast cancer and intrinsic subtypes of breast cancer determined by PAM50. In order to provide new molecular information to the already known markers for obtaining more sensitive and specific predictive models, the differences in gene expression of 730 genes implicated in cancer have been analyzed between the patients who have obtained pCR and those who have not from the Cohort_Discovery. This has been accomplished with Nanostring's nCounter® technology and the PanCancer Pathways gene panel. The p-value returned by the statistical test was adjusted by the FDR test (false discovery ratio) by the Benjamini-Hochberg method, with and adj. p-value <0.05 defining the DE genes. Once all the variables had been determined, a univariate analysis was performed using the Likelihood Ratio Test (LTR) as goodness of fit, to analyse how each of them were related to the pCR variable. For the determination of the predictive signatures, a multivariate analysis with double cross-validation was performed, internal validation with a LASSO model and external validation with Leave-one-out cross-validation (LOOCV). The variables have been divided into 4 sets, the **CLIN** data set, which includes the clinicopathological variables, the **GENEX** data set where only the information on the DE genes is included, and the **CLIN_GENEX** data set that combines the information of the DE genes with the clinical variables. The last set is **PAM50 genes** that only includes the variable of the intrinsic subtypes determined by PAM50. Predictive models have been determined from all these data sets and from all the cohorts. Once the models were defined, their evaluation was carried out by determining the area under the ROC curve (AUC). This process was repeated 1000 times to obtain more robust results. The same procedure was carried out in the validation cohort.

Results

The patients included in the Cohort_Discovery (n = 84), are mostly premenopausal (70%), with an average tumor size of 3.7 cm; in 60% of cases, lymph node involvement is positive, and 56% have histological tumor grade between 1 and 2. 60% is ER+, 50% PR+ and 25% HER2+. In 63% of the cases, Ki67 is less than 50%. The subtypes determined by IHC are: 40 Luminal (A + B), 21 HER2+ and 23 TN. As determined by PAM50, 9 Luminal A patients, 28 Luminal B patients, 22 HER2-enriched patients and 25 Basal-Like. Out of the 84 patients, 56 did not achieve pCR after treatment and 28 did.

The significantly differentially expressed genes (DE) between patients who have obtained pCR and those who have not, are 10: *FGFR4*, *CDC6*, *GNG4*, *MYCN*, *RASGRF1* and *IDH2*, more expressed in patients who obtain pCR, and *LEFTY1*, *FGF10*, *WNT3* and *STK11*, less expressed in patients who do not get pCR.

In the univariate analysis, the variables related to the pCR with statistical significance were the overexpression of HER2, the breast subtypes determined by IHC, the subtypes determined by PAM50 and the differential expression of the genes *CDC6*, *LEFTY1*, *FGFR4*, *MYCN*, *IDH2*, *STK11*, *FGF10*, *RASGRF1* and *WNT3*.

With multivariate predictive analysis, models with predictive capacity have been generated for the 4 data sets in the Cohort_D_Allsubtypes, Cohort_D_HER2- and Cohort_D_ER+ cohorts. The variables included in the models, according to the importance that the selection of characteristics by LASSO assigns to each one, are detailed below. The PAM50 model only includes the variable of the intrinsic breast subtypes defined by PAM50.

Regarding the variables selected by LASSO and the predictive capacity of the models, in cohort_D_Allsubtypes, the CLIN_Allsubtypes model (AUC 0.55) includes the HER2 overexpression variable; the GENEX_Allsubtypes model (AUC 0.79) includes the expression of *CDC6* genes, *FGFR4*, *LEFTY1*, *STK11* and *IDH2*. On the other hand, the combination of HER2 overexpression and clinical lymph node involvement, together with the information on the gene expression of the *FGFR4*, *LEFTY1*, *STK11*, *CDC6* and *IDH2*

genes, form the CLINGENEX_Allsubtypes predictive model (AUC 0.74). The PAM50 model in this cohort obtained an AUC of 0.51.

In cohort_D_HER2-, the CLIN_HER2- (AUC 0.6) model includes the variables histological tumor grade and tumor size, the GENEX_HER2- (AUC 0.83) model includes the expression of the *FGFR4*, *LEFTY1* and *STK11* genes. The combination of tumor size and clinical lymph node involvement, ER status plus information on the gene expression of the *FGFR4*, *LEFTY1* and *STK11* genes, form the CLINGENEX_HER2- (AUC 0.85) predictive model. The PAM50 model in this cohort obtained an AUC of 0.52.

In cohort_D_ER+, the CLIN_ER+ (AUC 0.71) model includes the variables clinical tumor size, histological tumor grade and clinical lymph node involvement, the GENEX_ER+ model (AUC 0.96) includes the expression of the *CDC6*, *FGFR4*, *LEFTY1*, *RASGRF1*, *WNT3* and *GNG4*. The combination of clinical lymph node involvement, histological tumor grade, tumor size, PR status together with the information on the gene expression of the *FGFR4*, *LEFTY1*, *WNT3*, *CDC6* and *GNG4* genes, form the CLINGENEX_ER+ predictive model (AUC 0.96). The PAM50 model in this cohort obtained an AUC of 0.76.

The ratios of the AUCs between the predictive models defined in this work and PAM50 give us an idea of the comparative advantage of the models. Among all the cohorts, the highest value of the ratio is found between the CLINGENEX_HER2- and PAM50 model (1.62) and the GENEX_HER2- and PAM50 model (1.58), in patients with HER2- tumors. In the cohort that includes all breast cancer subtypes, we observe the predictive advantage between the GENEX_Allsubtypes and PAM50 (1.54) and CLINGENEX_Allsubtypes and PAM50 (1.45) models. The models where only clinicopathological variables are included do not have great advantages with respect to the PAM50-based one.

Regarding the validation of the results, the improvement in terms of prediction that we have observed with the inclusion of the gene expression information in the models is maintained. It is worth highlighting the CLINGENEX_Allsubtypes model, in which the AUC remains at 0.74 in the validation cohort, and in the ER+ cohort, for which the predictive value of CLINGENEX_ER+ has an AUC of 0.76.

Discussion

In the context of predicting response to treatment with NAC, we propose that the information on the gene expression of the following DE genes from the panel of 730 genes analyzed among patients with pCR and non-pCR: *FGFR4*, *CDC6*, *GNG4*, *MYCN*, *RASGRF1*, *IDH2*, *STK11*, *WNT3*, *FGF10* and *LEFTY1*, provide greater predictive power and is reflected in the improvement in the AUCs of the models that include them, as is the case of CLINGENEX in the cohort of all subtypes or when it is limited to a ER+ population. Of these genes, we want to highlight the role of the *FGFR4* and *LEFTY1* genes since they are always chosen in the prediction algorithm in all the analysed cohorts. *FGFR4* is a membrane tyrosine kinase, and the aberrant activation of this protein has been associated with tumor progression and survival (22). *LEFTY1* is a tumor suppressor gene, related to the proliferation of progenitor cells of the mammary epithelium, and to stemness phenotypes (23).

In order to verify the real scope of the predictive improvement of the models that we propose with the pCR markers that currently exist, we designed a model with the clinicopathological variables (CLIN), and another one that only includes the variable PAM50 (1). In all cases, the inclusion or composition of the expression variable of the DE genes in the models increased the predictive power of pCR in relation to the other models based solely on the clinicopathological variables, which include PAM50, or PAM50 as the only variable. Therefore, these DE genes provided signatures based on the differences in gene expression in specific oncogenic processes between patients with and without pCR, which are also not mere surrogate markers of the tumor subtype as is the case with PAM50.

Tumors with ER positive status are the least responsive to treatment with NAC and, therefore, the most susceptible to the suboptimal use of the chemotherapeutic treatment. Due to this, identifying markers that better predict which patients will obtain pCR from those who will not, becomes even more vital in this type of tumors. Based on our results, the CLINGENEX_ER+ signature, with an AUC in our cohort of 0.96, and validated with an AUC of 0.76, is positioned as the model with the highest predictive

capacity of response to treatment with NAC, surpassing the current best predictor, which is PAM50, in patients with ER+ tumors.

Conclusions

The information of the genes with differential expression not only provides predictive power to the variables already known, but, without the need for the use of classic response markers, it is capable of differentiating the patients who are going to obtain pCR from those who do not with better predictive value. Especially in the case of ER+ patients where the percentage of patients who respond to NAC is minimal, the CLINGENEX_ER+ signature can identify with an AUC of 0.96 in our cohort those patients who will respond worse and could avoid the adverse effects of the treatment with NAC. The information contained in the signatures that include these predictive models has made it possible to formulate several hypotheses regarding the mechanisms of resistance to neoadjuvant chemotherapy. To test these hypotheses, it is necessary to further study the role of the DE genes in our cohort in the response to treatment with NAC, especially the *FGFR4* and *LEFTY1* genes, which in all cohorts are selected as an important part of the pCR prediction algorithm.

CHAPTER II: Resistance to treatment with NAC

The response to treatment with NAC is key to the prognosis of the patient after surgery. There are degrees of response ranging from pCR, which is the absence of tumor cells, to extensive residual disease, through a spectrum of intermediate RD values (1). The presence of RD after treatment with NAC is a fundamental prognostic factor. Reduction in tumor burden after treatment has been shown to be associated with increased disease-free survival (DFS) and overall survival (OS) (9) (10) (13) (3). With the idea of being able to classify the clinical result according to the characteristics of RD, biomarkers have been determined that provide an idea of the tumor status and this in turn has been related to a specific clinical result. The determination of Ki67 in the RD, the residual

cancer burden index (RCB) and the residual proliferative cancer burden (RPCB) are some of the markers used to assign levels in the RD (1).

Due to the aggressiveness of triple negative breast cancer (TN) when the cytotoxic effect of NAC does not entirely eliminate the tumor, TN BC is the type of breast cancer where RD has been most thoroughly studied at the molecular level. Indeed, in this subtype it has been found that NAC induces a series of chaotic acquisitions and losses of copy number alterations (CNA) in the genome of the cells that survive it. Amplifications in the *MCL1*, *MYC*, *KRAS*, *BCL9*, *EGFR*, *CDKN2A*, *BIRC2* and *BIRC3* genes are found in the RD stage when compared with the tumor before treatment (24). The expression of the *AR*, *GATA3*, *ESR2*, *GBX2*, *KRT16*, *MMP28* and *WNT11* genes in the RD of TN patients form a predictive model for relapse-free survival (RFS) in three years. Those TN patients who in RD overexpress genes such as *AR* or *GATA3* (related to Luminal phenotypes) will have an RFS of 76.9% compared to 25% of patients who overexpress genes related to Cancer Stem Cell markers (CSC) such as *KRT16*, *WNT11* or *MMP28* 87.

Beyond residual disease, deterministic metastatic factors are largely lacking after NAC treatment and a deeper understanding of their relationship to specific molecular profiles of the primary tumor and the RD. A more important concern than even the detection or prediction of RD is the development of the distant relapse, as it is responsible for 90% of breast cancer-related deaths (25). Currently, there are studies where several diagnostic markers of metastatic breast cancer have been established; however, no single predictor of metastasis from the RD stage has been determined.

Therefore, we consider important, and therefore have established as objectives in this chapter of the work, 1) the longitudinal study of a homogeneous cohort of NAC unresponsive breast cancer patients with paired samples of primary tumor (Pre_T1), RD (RD_T2) and metastasis (Mtx_T3) if possible) for the identification of new therapeutic targets in RD and 2) to determine the impact of treatment on the development of distant metastases.

Study population and methodology

A retrospective analysis has been carried out with patients under the following inclusion criteria: that they are adult women (older than 18 years) with breast cancer treated with NAC (Anthracyclines +/- Taxanes and targeted therapy with Trastuzumab in HER2+ patients) and non-responders to treatment with residual disease and metastases. That they had at least one sample of tumor tissue from both the primary tumor before treatment (Pre_T1) and the residual disease RD (RD_T2) and, if possible, tumor tissue material from the first metastatic niche (Mtx_T3).

All analyzes have been carried out on paraffin-embedded tissue samples. Response to treatment has been defined in the surgical biopsy after neoadjuvant treatment, according to the Miller & Payne criteria.

A differential expression analysis (DEA) was carried out with the nSolver™ tool using the expression values of the 730 genes of the PanCancer Pathways panel that was generated with the Nanostring nCounter® technology; the p-value was adjusted by the Benjamini-Hochberg method. The statistical limit to define a gene as DE among each comparative group is established at an adjusted p-value of 0.05, so that those genes whose expression is different with a statistical significance below 0.05 will be considered ED, regardless of the magnitude (Log2 FC) of the difference. In all comparisons, the smallest time has been taken as a reference.

Results

The patients included in this part of the study (n = 70 for the comparison Pre_T1 and RD_T2 and n = 22 for the comparison Pre_T1, RD_T2 and Mtx_T3) are mostly premenopausal (54-68%); they have a clinical tumor size of between 3.7 and 4 cm; most of them have positive clinical involvement (54-57%); the tumor grade between 1 and 2 (55-59%) predominates, the ER+ status is in 61 and 50% of patients, and the overexpression of HER2 in 17 and 27%, respectively. The distribution of PAM50 subtypes for the cohort where Pre_T1 is compared with RD_T2, was of 9 Luminal A, 26 Luminal B,

14 HER2enriched, and 19 Basal-Like. For the comparison Pre_T1, RD_T2 with Mtx_T3 the distribution was 5 Luminal A, 4 Luminal B, 6 HER2enriched and 5 Basal-Like.

In the comparison between the primary tumor and DR (Pre_T1vsRD_T2), we found 112 DE genes out of the 730 analysed genes. Given their statistical power and difference in the amount of change, it is worth highlighting the genes *FOS* ($p = 2.57E-23$, Log2 FC - 3.55), *NR4A1*, and *NR4A3* ($p = 5.4E-16$ and Log2 FC of -2.8 for *NR4A1*; $p = 4.4E-11$ and Log2 FC of -2.2 for *NR4A3*). These genes are downregulated in the primary tumor and show an increase in expression after treatment with NAC. *FOS* is part of the transcription factor AP1. *FOS* has been previously related to resistance to chemotherapeutic treatment (26) (27), consistently with the findings of our study. This gene has been associated with cellular adaptation to changes in the tumor environment in cell lines and the low expression of this gene has been associated with sensitivity to treatment (28). *NR4A1* and *NR4A3* are thyroid hormone receptors. The decrease in *NR4A1* has been related to an attenuation of the JNK-AP-1-Cyclin D1 pathway (29), which could also be related to the low values of *FOS* in the primary tumors of our cohort. *NR4A1* has also been shown to be a very potent activator of the TGF- β pathway and to contribute to tumor progression, epithelial-mesenchymal transition processes, and cell migration (30), which is consistent with the resistant phenotype of RD. On the other hand, *NR4A3* has tumor suppressor and pre-apoptotic functions (31).

Among the 112 DE genes, there are members of the Wnt family (*WNT7B*, Log2 FC 1.1; *WNT16*, Log2 FC -1.4; *WNT2*, Log2 FC 0.9; *WNT11*, Log2 FC -0.8), membrane receptors (*PRLR*, Log2 FC 1; *IL1R2*, Log2 FC -0.8; *LIFR*, Log2 FC -0.7; *LEPR*, Log2 FC -0.9; *CSF3R*, Log2 FC 0.8), phospholipases (*PLA2G5*, Log2 FC -0.8; *PLA2G2A*, Log2 FC -1.3; *PLA2G10*, Log2 FC 0.9; *PLA2G4E*, Log2 FC -1.5), voltage-dependent calcium channel subunits (*CACNG4*, Log2 FC 2; *CACNA2D3*, Log2 FC -0.9; *CACNG1*, Log2 FC 0.9; *CACNB3*, Log2 FC 0.8; *CACNA1G*, Log2 FC -0.7) and subunits of integrins (*ITGA2*, Log2 FC 0.8; *ITGB6*, Log2 FC 1.6; *ITGB4*, Log2 FC 0.8; *ITGA7*, Log2 FC -1.2).

On the other hand, the *PLAU* and *STAT* genes, which have previously been related to sensitivity to treatment with NAC when the tumor is compared before and after treatment

(32), are more expressed in our cohort in those patients that are sensitive to the treatment (Log2 FC 1.12, $p = 0.0008$ and Log2 FC 1.34, $p = 0.0004$, respectively) in coherence with the previous results. The expression of the *WNT11*, which has been reported as upregulated in those patients who do not respond satisfactorily to treatment with NAC in the same work (32), is also associated with resistance in our dataset (Log2 FC -0.8, $p = 0.038$).

In the comparison between residual disease and metastatic disease (RD_T2vsMtx_T3), we have identified 30 DE genes. Of these, three genes whose expression is increased in metastatic disease are of special importance: the *MAPK8IP2* gene (Log2 FC -2.8, $p = 3.98E-05$), the *IL22R* gene (Log2 FC -2.71, $p = 1.14E-04$) and the *LAMB4* (Log2 FC -2.5, $p = 0.0014$). The protein product of the *MAPK8IP2* gene is related to the increase in stemness, and dedifferentiation processes [30, p. 7]. In line with this, *MAPK8IP2* is more expressed in the metastatic tumor in our cohort. The *IL22RA2* gene product is a subunit of interleukin 22 and is related to the inflammatory response and regulation of tumorigenesis in the colon (RefSeq, 2013). The *LAMB4* gene product mediates the organization and migration of cells in tissues and, furthermore, has been associated with decreased survival of ER+ breast tumors (33). The overexpression of this gene in our metastatic samples could be due to the metastatization process where a reorganization of the cells in the new tissue occurs.

Finally, the comparison between the primary tumor and the metastatic disease (Pre_T1vsMtx_T3) yielded 11 DE genes. Like in the Pre_T1vsRD_T2 comparison, the *FOS* gene is positioned as the gene with the major statistical significance and the greatest amount of change between the primary tumor and the metastatic disease (Log2 FC -3.5, $p = 6.28E-06$).

Overlapping the comparisons in pairs, when the primary tumor is compared with RD and with metastasis (Pre_T1vsRD_T2 and Pre_T1vsMtx_T3), the *IFNA17*, *GRIN1*, *FOS*, *NR4A1*, *WNT16* and *CAMK2B* genes are DE in both analyzes, being upregulated in both transitions, from the primary tumor to metastasis, with the exception of the *CAMK2B* gene. The overlap of the DE genes of RD versus primary tumor and metastasis

(Pre_T1vsRD_T2 and RD_T2vsMtx_T3), yields *RAC3*, *FGF1*, *IBSP*, *WNT16*, *CAMK2B*, *ITGB6*, *MAPK8IP2*, *RASGRP1*, and *MLLT4* as the common DE genes. In this case, it is important to note that there is a decrease in the expression of DE genes in RD, both when compared with the primary tumor and when compared with metastasis, with the exception of *WNT16*, whose expression seems to be selected along the evolution of the tumor, increasing in RD with respect to the primary tumor and continuing the increment towards metastasis. Finally, the common DE genes in the metastatic tumor versus the primary tumor and RD (Pre_T1vsMtx_T3 and RD_T2vsMtx_T3) are *LAMB4*, *WNT16*, *NPM2* and *CAMK2B*; in both comparisons, their expression is increased in metastatic disease, except for the *CAMK2B* gene that, although it is more expressed in metastasis with respect to the primary tumor and residual disease, it is downregulated in the transition to RD.

The overlap of all the comparison unmasks two common genes: *WNT16* and *CAMK2B*. In the case of the *WNT16* gene, we have already commented on this possible positive selection of its expression after treatment with NAC, due to the progressive increase in gene expression between Pre_T1, RD_T2 and Mtx_T3. Therefore, it is interesting to study it as a candidate for a quantitative diagnostic marker of these three stages of the disease. In the case of *CAMK2B*, we see a decrease in gene expression after treatment that is ultimately recovered in the metastatic tumor.

Discussion

There are a series of DE genes between the primary tumor and the residual disease that have diverse functions; some are transcription factors, such as *FOS* and *JUN* that form the AP1 transcription factor complex related to the regulation of proliferation and apoptosis, and hormone receptors such as *NR4A1* and *NR4A3*; others are membrane receptors such as *PRLR*, *IL1R2*, *LIFR*, *LEPR* and *CSF3R*; there are also several DE members of the Wnt family (*WNT7B*, *WNT16*, *WNT2* and *WNT11*), genes related to voltage-dependent calcium channels (*CACNG4*, *CACNA2D3*, *CACNG1*, *CACNB3* and *CACNA1G*) and several types of integrins (*ITGA2*, *ITGB6*, *ITGB4* and *ITGA7*), among others. All those

changes that are positively selected after treatment with NAC provide resistance and characterize a population that has acquired the molecular mechanisms that are necessary to cope with cytotoxic treatment. Of the DE genes between the primary tumor and DR identified in this study, *PLAU*, *STAT1* and *WNT11* had already been described as genes with changes in their expression after neoadjuvant treatment in previous studies (1) (32). The expression profile of these genes correspond to the profile that we have obtained in our analyzes, supporting their role in the sensitivity to treatment with NAC (in the case of *PLAU* and *STAT1*) or in the resistance to it (in the case of *WNT11*). If we include the metastatic tumor in the comparison, it should be noted that *FOS*, *NR4A1*, *GRIN1* and *IFNA17* are found DE and in the same direction and magnitude in both comparisons, that is, less expressed in the primary tumor compared to DR and less expressed between the primary tumor with respect to metastasis, without significant differences between DR and metastasis. This indicates that, once its expression has decreased in the residual tumor, it remains constant and is not informative of the transition to metastasis. On the other hand, there is a series of genes such as *RAC3*, *FGF1*, *IBSP*, *ITGB6*, *MAPK8IP2*, *RASGRP1* and *MLLT4*, which are more overexpressed in the primary tumor and in metastatic disease than in RD, without significant differences between the primary tumor and metastatic tumor. In this case, we could assume that they are genes that will determine metastasis and that they may be related to the aggressiveness of the tumor from the start of treatment with NAC. Regarding the metastatic disease, we could define that the *LAMB4* and *NPM2* genes are biomarkers of distant relapse when compared with the primary tumor and with RD. The *WNT16* gene, which is associated with taxane-induced progression and chemoresistance (34), is positively selected after NAC, being this selection progressive; its expression increases from the primary tumor to the RD and from the RD to the metastatic tumor, which indicates a plausible selection of super clones. On the other hand, the *CAMK2B* gene, involved in calcium signaling, has been associated in previous studies with response to treatment with NAC (35). In our cohort, this gene is more expressed in the primary tumor than in RD after NAC, supporting the existing data. According to our longitudinal analysis, the gene expression of *CAMK2B* increases significantly even much more in

metastasis than in the primary tumor, where it has a higher expression than in RD. Therefore, it becomes a candidate for a metastasis biomarker.

Conclusions

Based on our results and the proposed objectives, a predictive signature has been identified based on the biological differences found between patients who obtain pCR from and those who do not, independently of the subtype, and which improves the current prediction tools. This improvement in the identification of responding versus non-responding patients is more notable in ER+ patients; This is especially relevant because these patients are the ones who least respond to NAC and therefore the ones who would benefit the most from a more specific discrimination.

In addition, both common and specific gene expression alterations have been identified in the primary tumor when compared with residual disease and the first metastatic niche. An inactivation of the gene expression of the *IFNA17*, *GRIN1*, *FOS*, *NRA41* and *WNT16* genes is shown in the primary tumor against residual disease and metastatic disease, where they are more expressed. The overexpression of these genes in RD could provide a series of new markers that need to be further investigated to propose new therapeutic targets in residual disease once treatment with NAC has not been effective. Regarding the metastasis, our results propose four DE genes between metastasis versus residual disease and primary tumor. *LAMB4* and *NPM2* as DE genes with higher expression only in metastasis, *CAMK2B* as a gene that is selected negatively after NAC in residual disease but that recovers its activity in metastatic tumor, and *WNT16* as a possible marker of unfavorable evolution of the tumor after NAC towards metastasis.

Impact

The results of this doctoral thesis represent an advance in the molecular knowledge of the different moments of breast cancer treated with NAC from the primary tumor to the distant relapse. Predictive models of pCR, as well as the molecules that differentially

define RD and metastasis are the basis for the implementation of precision medicine and the discovery of new therapeutic targets for resistance in this highly prevalent disease, and whose relapse involves 90% of deaths related to it.



UNIVERSIDAD
DE MÁLAGA

Lista de abreviaturas

AIC	<i>Akaike Information Criteria</i>
AED	Análisis de expresión diferencial
ASR	Ratio de edad estandarizado (<i>age-standardized rates</i>)
AUC	Área bajo la curva (<i>area under the curve</i>)
BIC	<i>Bayesian Information Criteria</i>
BL1	Basal-Like 1
BL2	Basal-Like 2
CISH	Hibridación in situ cromogénica
CK5/6	Citoqueratinas 5/6
CM	Cáncer de mama
CNAs	Alteraciones en el número de copias (<i>copy number alterations</i>)
cT	Tamaño tumoral clínico
DE	Diferencialmente expresados
EGFR	Receptor del factor de crecimiento epidérmico
EMT	Transición Epitelio-Mesénquima (<i>epithelial-mesenchymal transition</i>)
ER	Receptor de Estrógenos (<i>estrogen receptor</i>)
FFPE	Muestras incluidas en bloques de parafina (<i>formalin-fixed paraffin-embedded</i>)
FISH	Hibridación in situ fluorescente
H&E	Hematoxilina y Eosina
HER2	Factor de crecimiento epidérmico humano 2 (<i>human epidermal growth factor 2</i>)
HR	<i>Hazard ratio</i>
IHQ	Inmunoquímicos
ILRD	Intervalo libre de recurrencia a distancia
IM	<i>Immunomodulatory</i>
IntClust	<i>Integrative Cluster</i>
LAR	Luminal androgen receptor
Log2 FC	logaritmo en base 2 del <i>Fold Change</i>

LOOCV	<i>Leave-one out cross-validation</i>
LRT	<i>Likelihood Ratio Test</i>
M	<i>Mesenchymal</i>
MCE	Muerte con enfermedad
MSL	<i>Mesenchymal stem-like</i>
NAC	Quimioterapia neoadyuvante (<i>Neoadjuvant chemotherapy</i>)
OMS	Organización mundial de la salud
OR	Razón de momios (<i>Odds Ratio</i>)
pCR	Respuesta completa patológica (<i>pathological complete response</i>)
PR	Receptor de progesterona (<i>progesterone receptor</i>)
RCB	Carga del tumor residual (<i>residual cancer burden</i>)
RD	Enfermedad residual (<i>residual disease</i>)
RL	Regresión logística
RPCB	Carga residual proliferativa del cáncer (<i>Residual Proliferative Cancer Burden</i>)
RUV	<i>Removing unwanted variation</i>
SG	Supervivencia global
SLE	Supervivencia libre de enfermedad
SLR	Supervivencia libre de recaída
SLRD	Supervivencia Libre de Recaída a Distancia
SNP	Polimorfismos de nucleótido único (<i>single nucleotide polymorphism</i>)
TCGA	<i>The Cancer Genoma Atlas</i>
TILs	Linfocitos infiltrantes en tumor (<i>Tumor infiltrating lymphocytes</i>)
TN	Triples negativos
TRAR	<i>Trastuzumab Risk model</i>

Índice

1. INTRODUCCIÓN GENERAL	1
1.1. EPIDEMIOLOGÍA DEL CÁNCER DE MAMA.....	3
1.2. HETEROGENEIDAD EN EL CÁNCER DE MAMA: SUBTIPOS Y CLASIFICACIONES.....	4
1.2.1. <i>Grado y clasificación histológica</i>	4
1.2.2. <i>Subtipos Inmunohistoquímicos</i>	6
1.2.3. <i>Subtipos intrínsecos o moleculares</i>	7
1.3. QUIMIOTERAPIA NEOADYUVANTE EN EL CÁNCER DE MAMA	8
1.3.1. <i>pCR y enfermedad residual</i>	8
1.3.2. <i>Factores predictivos de pCR</i>	10
1.3.2.1. Factores clinicopatológicos predictivos de pCR.....	10
1.3.2.2. Factores moleculares predictivos de pCR.....	11
1.3.2.2.1. Subtipos intrínsecos determinados por PAM50.....	11
1.3.2.2.2. Otros predictores moleculares de pCR	12
1.4. ENFERMEDAD RESIDUAL EN CÁNCER DE MAMA TRAS EL TRATAMIENTO NEOADYUVANTE.....	13
1.4.1. <i>Determinación del Ki67 después del tratamiento Neoadyuvante</i>	14
1.4.2. <i>Índice de carga residual de cáncer</i>	14
1.4.3. <i>Carga residual proliferativa del cáncer</i>	16
1.4.4. <i>Dinamismo molecular de la RD después del tratamiento con NAC</i>	16
1.4.5. <i>Perfil transcriptómico y genómico de la RD tras la NAC</i>	17
1.4.5.1. Alteraciones moleculares en la RD de tumores TN	18
1.5. TUMOR PRIMARIO, ENFERMEDAD RESIDUAL Y METÁSTASIS	19
2. HIPÓTESIS Y OBJETIVOS	21
2.1. HIPÓTESIS	23
2.2. OBJETIVOS.....	23
3. MATERIAL Y MÉTODOS GENERALES	25
3.1. PACIENTES Y TRATAMIENTOS	27
3.2. EVALUACIÓN HISTOLÓGICA, pCR Y SUBTIPOS IHQ DE MAMA.....	29
3.3. ESTUDIO DE LA EXPRESIÓN GÉNICA.....	31
3.3.1. <i>Aislamiento del mRNA desde el tejido incluido en parafina</i>	31
3.3.2. <i>Determinación de la expresión génica con la tecnología nCounter Dx Analysis System (Nanostring technologies)</i>	31
3.3.2.1. Subtipos moleculares de mama determinados por PAM50	33
3.3.2.2. Panel <i>PanCancer pathways</i> : Análisis de expresión diferencial, correlación entre genes diferencialmente expresados y enriquecimiento en rutas biológicas.....	34
3.3.2.2.1. Control de calidad y normalización de los datos.....	36
3.3.2.2.2. Análisis de expresión diferencial en T1 entre pCR y no-pCR (Capítulo I)....	37
3.3.2.2.3. Análisis de expresión diferencial en el análisis longitudinal (capítulo II)....	37
3.3.2.2.4. Enriquecimiento en rutas biológicas.....	38

3.3.3.	<i>Predicción de pCR en T1</i>	39
3.3.3.1.	Análisis Univariante	40
3.3.3.2.	Análisis multivariante.....	40
3.4.	ESTUDIO DE LAS ALTERACIONES GENÉTICAS	44
3.4.1.	<i>Aislamiento del DNA desde el tejido embebido en parafina.</i>	44
3.4.2.	<i>Genes y secuenciación.</i>	44
3.4.3.	<i>Análisis bioinformático de las variantes encontradas a nivel de DNA:</i> <i>Mutaciones</i>	47
3.5.	ANÁLISIS DE SUPERVIVENCIA.....	48
3.6.	LIMITACIONES DE LOS ESTUDIOS	49
4.	CAPÍTULO I:	51
4.1.	INTRODUCCIÓN Y OBJETIVOS.....	53
4.2.	POBLACIÓN DE ESTUDIO Y METODOLOGÍA.....	55
4.3.	MUTACIONES ANTES DEL TRATAMIENTO CON NAC (DNA-COHORTE).....	63
4.4.	RESULTADOS.....	64
4.4.1.	<i>Expresión génica (RNA-Cohorte)</i>	64
4.4.1.1.	Características clínico-patológicas	64
4.4.1.2.	Panel <i>PanCancer pathways</i> en la cohorte <i>Discovery</i> . Análisis de expresión diferencial (AED), correlación entre genes diferencialmente expresados y enriquecimiento en rutas biológicas.....	70
4.4.1.3.	Predicción de pCR en Cohorte <i>Discovery</i>	79
4.4.1.3.1.	Predicción de pCR en la Cohorte_D_Allsubtypes	81
4.4.1.3.2.	Predicción de pCR en la Cohorte_D_HER2-	85
4.4.1.3.3.	Predicción de pCR en la Cohorte_D_ER+	89
4.4.1.3.4.	Mejora de los modelos predictivos frente a PAM50 como predictor individual	92
4.4.1.3.5.	Validación de los modelos predictivos.....	94
4.4.2.	<i>Alteraciones genéticas (DNA-Cohorte)</i>	99
4.4.2.1.	Características clinicopatológicas	99
4.4.2.2.	Descriptiva de las mutaciones encontradas en T1	101
4.4.2.3.	Alteraciones en T1 según la respuesta a la NAC.....	103
4.5.	DISCUSIÓN	105
4.6.	CONCLUSIONES	110
5.	CAPÍTULO II:	113
5.1.	INTRODUCCIÓN Y OBJETIVOS.....	115
5.1.1.	<i>Marcadores moleculares de RD</i>	116
5.1.2.	<i>Dinamismo en los biomarcadores ER, PR y HER2 tras la NAC</i>	116
5.2.	POBLACIÓN DE ESTUDIO Y METODOLOGÍA.....	118
5.2.1.	<i>Pacientes y muestras</i>	118
5.2.2.	<i>Inmunohistoquímica, subtipos por PAM50 y determinación de la RD</i> ...	119
5.2.3.	<i>Determinación de la expresión génica</i>	120
5.2.3.1.	Panel PanCancer Pathways	120

5.3.	RESULTADOS.....	120
5.3.1.	<i>Características clínico-patológicas.....</i>	<i>120</i>
5.3.2.	<i>AED del panel PanCancer Pathways en Pre_T1, RD_T2 y Mtx_T3.</i>	<i>124</i>
5.3.2.1.	Tumor primario (Pre_T1) versus enfermedad residual (RD_T2).	124
5.3.2.2.	Enfermedad residual (RD_T2) versus metástasis (Mtx_T3).	132
5.3.2.3.	Tumor primario (Pre_T1) versus metástasis (Mtx_T3).....	136
5.3.2.4.	Tumor primario, enfermedad residual y metástasis	139
5.4.	DISCUSIÓN	142
5.5.	CONCLUSIONES	144
6.	CONCLUSIONES	147
7.	CONCLUSIONS.....	151
8.	REFERENCIAS.....	155



UNIVERSIDAD
DE MÁLAGA

Lista de figuras

FIGURA 1. DATOS DE LA FUENTE <i>GLOBAL CANCER STATISTICS 2020</i> : GLOBOCAN.	4
FIGURA 2. PRINCIPALES TIPOS Y GRADOS HISTOLÓGICOS EN EL CÁNCER DE MAMA. A) TIPOS HISTOLÓGICOS: DESDE IZQUIERDA ARRIBA HASTA ABAJO DERECHA: CÁNCER DE MAMA DUCTAL INVASIVO, MEDULAR, TUBULAR, CRIBIFORME, MUCINOSO Y METAPLÁSICO. B) GRADOS HISTOLÓGICOS: (A) TUMOR DE MAMA BIEN DIFERENCIADO (GRADO 1), (B) TUMOR DE MAMA MODERADAMENTE DIFERENCIADO (GRADO 2) Y (C) POBREMENTE DIFERENCIADO (GRADO 3). FIGURAS OBTENIDAS DEL ATLAS DE GENÉTICA Y CITOGENÉTICA EN ONCOLOGÍA Y HEMATOLOGÍA.....	5
FIGURA 3. SUBTIPOS DE CÁNCER DE MAMA CON TINCCIONES INMUNOHISTOQUÍMICAS DE LOS MARCADORES ER, PR Y HER2. LA HISTOLOGÍA DEL CÁNCER SE HA DETERMINADO USANDO UNA TINCIÓN CON HEMATOXILINA Y EOSINA. LA IMAGEN PROCEDE DEL ARTÍCULO DE RIVENBARK Y COLABORADORES, 2013.....	6
FIGURA 4. DISTRIBUCIÓN DE LAS PACIENTES EN LOS CAPÍTULO I Y II.	28
FIGURA 5. FLUJO DE TRABAJO DEL SISTEMA nCOUNTER®. 1) CADA RNA MENSAJERO HIBRIDA CON UNA SONDA REPORTER (QUE LLEVA LA SEÑAL) Y CON UNA SONDA DE CAPTURA (QUE SE ENCUENTRA INMOVILIZADA EN EL CARTUCHO). 2) LAS MUESTRAS SON CARGADAS EN LA ESTACIÓN DE PREPARACIÓN O PREPSTATION, DONDE EL EXCESO DE SONDAS SE RETIRA Y EL COMPLEJO RNA MENSAJERO–SONDA REPORTER–SONDA DE CAPTURA ES INMOVILIZADA Y ALINEADA EN LA SUPERFICIE DEL CARTUCHO. 3) CON EL ANALIZADOR DIGITAL O DIGITAL ANALYZER, SE RECOGEN LOS DATOS DE NUESTRAS MUESTRAS; LOS CÓDIGOS DE COLOR DE LA SUPERFICIE DEL CARTUCHO SON CONTADAS Y ASIGNADAS A CADA DIANA. LA IMAGEN HA SIDO TOMADA DE LA WEB HTTPS://WWW.NANOSTRING.COM/SCIENTIFIC-CONTENT/TECHNOLOGY-OVERVIEW/NCOUNTER-TECHNOLOGY	32
FIGURA 6. GENES INCLUIDOS EN LA FIRMA PAM50. EN LA PARTE SUPERIOR DEL <i>CIRCUS PLOT</i> SE MUESTRAN LOS GENES QUE SE INCLUYEN EL PANEL PAM50. EN LA PARTE INFERIOR SE MUESTRAN LOS PROCESOS BIOLÓGICOS A LOS QUE PERTENECEN DICHS GENES.	33
FIGURA 7. IMAGEN DE LA HERRAMIENTA nSOLVER™ CON LOS PARÁMETROS EN LOS CONTROLES DE CALIDAD Y LOS VALORES LÍMITES PARA QUE LAS MUESTRAS SEAN INCLUIDAS EN LOS POSTERIORES ANÁLISIS. .	36
FIGURA 8. VÍAS INCLUIDAS EN EL PANEL <i>PANCANCER PATHWAYS</i> DE nCOUNTER®. LOS CÍRCULOS REPRESENTAN DE CADA UNA DE LAS 13 VÍAS CANÓNICAS EL NÚMERO DE GENES SELECCIONADOS PARA EL PANEL EN RELACIÓN CON EL NÚMERO TOTAL DE GENES CONOCIDOS PARA CADA VÍA IDENTIFICADO	

POR KEGG. LAS LÍNEAS HAN SIDO DIBUJADAS PARA MOSTRAR LA RELACIÓN DE LOS GENES QUE PERTENECEN A DIFERENTES VÍAS, CON EL GROSOR DE LA LÍNEA SE DETERMINA EL NÚMERO DE GENES COMPARTIDOS..... 39

FIGURA 9. FLUJO DE TRABAJO PARA LA DETERMINACIÓN DE LOS MODELOS PREDICTIVOS. CON LOS SETS DE DATOS CLIN, GENEX Y CLIN_GENEX SE HAN DETERMINADO LAS VARIABLES MÁS IMPORTANTES PARA LA DETERMINACIÓN DE PCR EN LAS PACIENTES DE LAS COHORTES_ *DISCOVERY* CON UNA DOBLE VALIDACIÓN CRUZADA, UNA VALIDACIÓN INTERNA CON LASSO Y UNA VALIDACIÓN EXTERNA POR LOOCV. PARALELAMENTE SE HA INCLUIDO UN MODELO SOLO CON LA VARIABLE PAM50. DE NUEVO, LASSO HA SIDO EL MÉTODO USADO PARA EL ENTRENAMIENTO DE LOS MODELOS Y LA MÉTRICA PARA SU ESTIMACIÓN, AUC. UNA VEZ LAS FIRMAS HAN SIDO GENERADAS, HAN SIDO PROBADAS EN LAS DIFERENTES COHORTES TANTO EN LAS COHORTES_ *DISCOVERY* COMO EN LAS COHORTES_ *VALIDATION* USANDO LA MÉTRICA DE AUC PARA SU ESTIMACIÓN. EL MODELO PAM50 SOLO HA SIDO PROBADO EN LAS COHORTES_ *DISCOVERY*..... 43

FIGURA 10. FLUJO DE TRABAJO PARA LA SECUENCIACIÓN DE LOS GENES DIANA. A. LOS FRAGMENTOS DESNATURALIZADOS DE LAS LIBRERÍAS FLUYEN A TRAVÉS DE LA *FLOWCELL* UNIÉNDOSE A LOS ADAPTADORES QUE HAY PRESENTE EN LA SUPERFICIE, QUE SON COMPLEMENTARIOS A LOS ADAPTADORES DE NUESTRAS MUESTRAS. LOS FRAGMENTOS COMPLEMENTARIOS SON EXTENDIDOS Y AMPLIFICADOS CON UNA PCR–PUENTE, TRAS ESTO SERÁN DESNATURALIZADOS FORMANDO CLÚSTERES DESDE UN MISMO FRAGMENTO. B. DICHOS FRAGMENTOS SON SECUENCIADOS UTILIZANDO NUCLEÓTIDOS TERMINADORES REVERSIBLES. LOS PARES DE BASES SE IDENTIFICAN DESPUÉS DE LA EXCITACIÓN DEL LÁSER Y LA DETECCIÓN DE LA FLUORESCENCIA. C. LOS DATOS BRUTOS DE DESMULTIPLEXAN EN LIBRERÍAS INDIVIDUALES Y SE EVALÚA LA CALIDAD. SE ELIMINAN LAS LECTURAS DE LOS ADAPTADORES PARA ELIMINAR EL RUIDO TÉCNICO..... 46

FIGURA 11. DISTRIBUCIÓN DEL NÚMERO DE PACIENTES DE LA RNA–COHORTE SEGÚN SI HAN SIDO USADOS PARA EL DESCUBRIMIENTO DE LOS MODELOS PREDICTIVOS DE PCR TRAS EL TRATAMIENTO CON NAC O PARA LA VALIDACIÓN DE ELLOS..... 57

FIGURA 12. FLUJO DE TRABAJO PARA LA DETERMINACIÓN DE LOS MODELOS PREDICTIVOS. CON LOS SETS DE DATOS CLIN, GENEX Y CLIN_GENEX SE HAN DETERMINADO LAS VARIABLES MÁS IMPORTANTES PARA LA DETERMINACIÓN DE PCR EN LAS PACIENTES DE LAS COHORTES_ *DISCOVERY* CON UNA DOBLE VALIDACIÓN CRUZADA, UNA VALIDACIÓN INTERNA CON LASSO Y UNA VALIDACIÓN EXTERNA POR

LOOCV. PARALELAMENTE SE HA INCLUIDO UN MODELO SOLO CON LA VARIABLE PAM50. DE NUEVO, LASSO HA SIDO EL MÉTODO USADO PARA EL ENTRENAMIENTO DE LOS MODELOS Y LA MÉTRICA PARA SU ESTIMACIÓN, AUC. UNA VEZ LAS FIRMAS HAN SIDO GENERADAS, HAN SIDO PROBADAS EN LAS DIFERENTES COHORTES TANTO EN LAS COHORTES_ *DISCOVERY* COMO EN LAS COHORTES_ *VALIDATION* USANDO LA MÉTRICA DE AUC PARA SU ESTIMACIÓN. EL MODELO PAM50 SOLO HA SIDO PROBADO EN LAS COHORTES_ *DISCOVERY*..... 62

FIGURA 13. DIAGRAMAS DE CAJAS. SE REPRESENTAN CADA UNO DE LOS GENES DE CON UN DIAGRAMA DE CAJAS DONDE SE MUESTRA EL NÚMERO DE *COUNTS* NORMALIZADOS PROCEDENTES DEL ANÁLISIS DE LOS RESULTADOS OBTENIDOS (EJE DE ORDENADAS) FRENTE A LA SITUACIÓN DE ESTUDIO, PCR SÍ O NO (EJE DE ABCISAS). 72

FIGURA 14. EN EL *VOLCANO PLOT* SE REPRESENTAN TODOS LOS GENES SEGÚN EL P-VALOR AJUSTADO OBTENIDO EN EL AED (EJE DE ORDENADAS) Y EL VALOR DEL LOG2 FC (EJE DE ABCISAS). CON LAS LÍNEAS PUNTEADAS SE REPRESENTAN DIFERENTES NIVELES DE P-VALOR Y LA ZONA DEL VOLCANO DONDE SE POSICIONE EL GEN CORRESPONDERÁ CON UN P-VALOR CONCRETO Y UN VALOR DE FC DETERMINADO. AQUELLOS GENES EN LA ZONA IZQUIERDA DESDE EL VALOR 0 DEL EJE DE ABCISAS, SERÁN AQUELLOS GENES CON MENOR EXPRESIÓN EN LAS PACIENTES QUE OBTIENEN PCR, EN LA PARTE DERECHA ESTARÁN LOS GENES QUE SE EXPRESAN AL ALZA. 72

FIGURA 15. *CIRCOS PLOT*. SE REPRESENTA LOS 10 GENES DE ENLAZADOS A LAS RUTAS QUE PERTENECEN CON COLORES DIFERENTES. LOS GENES, A SU VEZ, SE LE ASIGNA UN COLOR QUE VA DESDE EL ROJO HASTA EL AZUL SEGÚN EL LOG2 FC QUE PRESENTEN CON RESPECTO A LA PCR. 76

FIGURA 16. EN LA MATRIZ SE MUESTRAN LAS CORRELACIONES ENTRE LOS 10 GENES DE. EL COLOR DE LOS NOMBRES DE LOS GENES DEPENDE DEL LOG2 FC OBTENIDO EN EL AED (EL COLOR ROJO PARA LOS GENES EXPRESADOS AL ALZA EN LAS PACIENTES CON PCR Y EN AZUL AQUELLOS EXPRESADOS A LA BAJA). EN LA BARRA LATERAL VEMOS UNA GRADACIÓN DE COLOR DESDE ROJO HASTA AZUL QUE CORRESPONDE AL VALOR *r* DE LA CORRELACIÓN DE PEARSON. DESDE EL ROJO MÁS INTENSO (CORRELACIÓN POSITIVA PERFECTA, $r=1$) HASTA EL AZUL MÁS OSCURO (CORRELACIÓN NEGATIVA PERFECTA, $r=-1$). ESTO MISMO SE HA REPRESENTADO EN CADA UNA DE LAS INTERACCIONES DENTRO DE LA MATRIZ. AQUELLAS INTERACCIONES QUE ESTADÍSTICAMENTE SIGNIFICATIVAS SE HAN IDENTIFICA CON UN ASTERISCO. . 77

FIGURA 17. A. DIAGRAMA DE BARRAS. REPRESENTACIÓN DE TODOS LAS VÍAS O RUTAS BIOLÓGICAS QUE INCLUIMOS EN EL ANÁLISIS FUNCIONAL. LA LÍNEA PUNTEADA DETERMINA EL P-VALOR POR DEBAJO DE

0.05. B, C Y D. DIAGRAMAS DE CAJAS. SE REPRESENTAN LAS TRES RUTAS ESTADÍSTICAMENTE SIGNIFICATIVAS SEGÚN SU COMPORTAMIENTO ENTRE LAS PACIENTES QUE OBTIENEN PCR Y LAS QUE NO, *CHROMATIN MODIFICATION*, *JAK-STAT* Y *TRANSCRIPTIONAL MISREGULATION PATHWAYS*. 78

FIGURA 18. REPRESENTACIÓN DE LA IMPORTANCIA DE LAS VARIABLES DENTRO DE CADA UNO DE LOS MODELOS PREDICTIVO DE PCR DE LA COHORTE_D_ *ALLSUBTYPES*. A. DIAGRAMA DE BARRAS DONDE SE MUESTRAN LAS VARIABLES CON MÁS PESO DENTRO DE LOS TRES SETS DE DATOS (CLIN, GENEX, CLINGENEX). EN EL EJE DE ORDENADAS SE IDENTIFICAN LAS VARIABLES EN ORDEN SEGÚN LA IMPORTANCIA DENTRO DEL MODELO, EN EL EJE DE ABCISA SE MUESTRA EL PORCENTAJE DEL NÚMERO DE VECES QUE LA VARIABLE SELECCIONADA POR LASSO OBTIENE UN COEFICIENTE 0. B. *CROSSBARS*. SE IDENTIFICA EL COEFICIENTE DE CADA VARIABLE. AL HABERSE ITERADO MIL VECES, SE IDENTIFICA EL INTERVALO DE CONFIANZA DEL 95%. EL PUNTO REPRESENTA LA MEDIA Y LA LÍNEA REPRESENTA EL INTERVALO. 82

FIGURA 19. A. CURVAS ROC DE LOS MODELOS PREDICTIVOS DE PCR DE LA COHORTE_D_ *ALLSCUBTYPES*. B. LOS *VIOLIN PLOTS* REPRESENTAN LA DISTRIBUCIÓN DEL AUC OBTENIDO DURANTE LAS MIL ITERACIONES. 84

FIGURA 20. REPRESENTACIÓN DE LA IMPORTANCIA DE LAS VARIABLES DENTRO DE CADA UNO DE LOS MODELOS PREDICTIVO DE PCR DE LA COHORTE_D_ *HER2-*. A. DIAGRAMA DE BARRAS DONDE SE MUESTRAN LAS VARIABLES CON MÁS PESO DENTRO DE LOS TRES SETS DE DATOS (CLIN, GENEX, CLINGENEX). EN EL EJE DE ORDENADAS SE IDENTIFICAN LAS VARIABLES EN ORDEN SEGÚN LA IMPORTANCIA DENTRO DEL MODELO, EN EL EJE DE ABCISA SE MUESTRA EL PORCENTAJE DEL NÚMERO DE VECES QUE LA VARIABLE SELECCIONADA POR LASSO OBTIENE UN COEFICIENTE 0. B. *CROSSBARS*. SE IDENTIFICA EL COEFICIENTE DE CADA VARIABLE. AL HABERSE ITERADO MIL VECES, SE IDENTIFICA EL INTERVALO DE CONFIANZA DEL 95%. EL PUNTO REPRESENTA LA MEDIA Y LA LÍNEA REPRESENTA EL INTERVALO. 86

FIGURA 21. A. CURVAS ROC DE LOS MODELOS PREDICTIVOS DE PCR DE LA COHORTE_D_ *HER2-*. B. LOS *VIOLIN PLOTS* REPRESENTAN LA DISTRIBUCIÓN DEL AUC OBTENIDO DURANTE LAS MIL ITERACIONES. 88

FIGURA 22. REPRESENTACIÓN DE LA IMPORTANCIA DE LAS VARIABLES DENTRO DE CADA UNO DE LOS MODELOS PREDICTIVO DE PCR DE LA COHORTE_D_ *ER+*. A. DIAGRAMA DE BARRAS DONDE SE MUESTRAN LAS VARIABLES CON MÁS PESO DENTRO DE LOS TRES SETS DE DATOS (CLIN, GENEX,

CLINGENEX). EN EL EJE DE ORDENADAS SE IDENTIFICAN LAS VARIABLES EN ORDEN SEGÚN LA IMPORTANCIA DENTRO DEL MODELO, EN EL EJE DE ABCISA SE MUESTRA EL PORCENTAJE DEL NÚMERO DE VECES QUE LA VARIABLE SELECCIONADA POR LASSO OBTIENE UN COEFICIENTE 0. B. *CROSSBARS*. SE IDENTIFICA EL COEFICIENTE DE CADA VARIABLE. AL HABERSE ITERADO MIL VECES, SE IDENTIFICA EL INTERVALO DE CONFIANZA DEL 95%. EL PUNTO REPRESENTA LA MEDIA Y LA LÍNEA REPRESENTA EL INTERVALO. 90

FIGURA 23. A. CURVAS ROC DE LOS MODELOS PREDICTIVOS DE PCR DE LA COHORTE_D_ER+. B. LOS *VIOLIN PLOTS* REPRESENTAN LA DISTRIBUCIÓN DEL AUC OBTENIDO DURANTE LAS MIL ITERACIONES. 91

FIGURA 24. *VIOLIN PLOTS*. SE REPRESENTA LA RATIO DEL AUC DE LOS MODELOS DEFINIDOS EN LAS DIFERENTES COHORTES VERSUS EL MODELO QUE INCLUYE SOLO LA VARIABLE PAM50. VALORES POR ENCIMA DE 1 INDICAN UNA MEJORÍA DE LOS MODELOS PROPUESTOS FRENTE AL MODELO PAM50. 93

FIGURA 25. VALIDACIÓN DE LOS MODELOS PREDICTIVOS CLIN_ *ALLSUBTYPES*/HER2-/ER+, GENEX_ *ALLSUBTYPES*/HER2-/ER+ Y CLINGENEX_ *ALLSUBTYPES*/HER2-/ER+ EN LAS COHORTES_ *VALIDATION*. A. CURVAS ROC DE LOS MODELOS PREDICTIVOS DE PCR EN LAS COHORTES_ *VALIDATION*. B. LOS *VIOLIN PLOTS* REPRESENTAN LA DISTRIBUCIÓN DEL AUC OBTENIDO DURANTE LAS MIL ITERACIONES DE CADA UNO DE LOS MODELOS EN LAS DIFERENTES COHORTES_ *VALIDATION*. 95

FIGURA 26. CURVA DE SUPERVIVENCIA DE KAPLAN-MEIER DE LAS PACIENTES INCLUIDAS EN ESTE ESTUDIO TENIENDO COMO EVENTO LA MUERTE CON ENFERMEDAD (MCE), EN ESTE CASO CÁNCER DE MAMA O METÁSTASIS DERIVADO DE ÉL. 97

FIGURA 27. SE MUESTRAN LAS CURVAS DE SUPERVIVENCIA KAPLAN-MEIER DE LAS VARIABLES QUE TIENEN UNA RELACIÓN ESTADÍSTICAMENTE SIGNIFICATIVA CON EL EVENTO MCE. LA COMPARACIÓN ENTRE CATEGORÍAS DENTRO DE LA MISMA VARIABLE SE HA REALIZADO CON UNA PRUEBA NO PARAMÉTRICO (LOG RANK). A. COMPARACIÓN ENTRE LA CATEGORÍA ER- Y ER+ DE LA VARIABLE ER. B. CURVAS DE LAS 4 CATEGORÍAS DENTRO DEL SUBTIPOS DE PAM50. C. DIFERENCIA ENTRE LAS CURVAS DE SUPERVIVENCIA ENTRE UN PORCENTAJE DE Ki67 MAYOR Y MENOR DE 50. D Y E. LA EXPRESIÓN DE LOS GENES FGF10 Y IDH2 SE HA DIVIDO EN 3 CUARTILES PARA DETERMINAR LAS DIFERENCIAS ENTRE SUS CURVAS DE SUPERVIVENCIA. 98

FIGURA 28. PORCENTAJE DE LOS GENES MUTADOS. 102



FIGURA 29. ONCOPLOTS. SE IDENTIFICAN LOS GENES EN LOS QUE SE HAN ENCONTRADO AL MENOS UNA MUTACIÓN Y LAS PACIENTES DE LA DNA-COHORTE. EL PORCENTAJE MOSTRADO EN LA DERECHA DE LA MATRIZ SE CORRESPONDE CON EL PORCENTAJE DE MUJERES CON MUTACIONES QUE TIENEN ESE GEN CONCRETO MUTADO. EL DIAGRAMA DE BARRAS EN EL LATERAL DERECHO ESPECIFICA EL TIPO DE MUTACIÓN ENCONTRADO DENTRO DE CADA GEN. EL DIAGRAMA DE BARRAS EN LA PARTE SUPERIOR DE LA MATRIZ IDENTIFICA EL TIPO DE MUTACIÓN ENCONTRADA DENTRO DE CADA PACIENTE. A. RESULTADOS DE TODAS LAS PACIENTES INCLUIDAS EN EL ESTUDIO. B. RESULTADOS DE LAS PACIENTES QUE NO OBTIENEN RESPUESTA A LA NAC. C. RESULTADOS DE LAS PACIENTES QUE OBTIENEN RESPUESTA A LA NAC. 104

FIGURA 30. ESQUEMA DE LA COHORTE DEL CAPÍTULO II. 118

FIGURA 31. EN EL *VOLCANO PLOT* SE REPRESENTAN TODOS LOS GENES SEGÚN EL P-VALOR AJUSTADO OBTENIDO EN EL AED (EJE DE ORDENADAS) Y EL VALOR DEL LOG₂ FC (EJE DE ABCISAS). CON LAS LÍNEAS PUNTEADAS SE REPRESENTAN DIFERENTES NIVELES DE P-VALOR Y LA ZONA DEL *VOLCANO* DONDE SE POSICIONE EL GEN CORRESPONDERÁ CON UN P-VALOR CONCRETO Y UN VALOR DE FC DETERMINADO. EN ESTE CASO, LOS GENES QUE QUEDEN A LA DERECHA DEL PUNTO 0 DEL EJE DE ABCISAS SERÁN LOS GENES SOBREEXPRESADOS EN EL PRE_T1 Y LOS QUE QUEDEN EN EL LADO IZQUIERDO SERÁN LOS SOBREEXPRESADOS EN EL RD_T2. 129

FIGURA 32. EN EL *VOLCANO PLOT* SE REPRESENTAN TODOS LOS GENES SEGÚN EL P-VALOR AJUSTADO OBTENIDO EN EL AED (EJE DE ORDENADAS) Y EL VALOR DEL LOG₂ FC (EJE DE ABCISAS). CON LAS LÍNEAS PUNTEADAS SE REPRESENTAN DIFERENTES NIVELES DE P-VALOR Y LA ZONA DEL *VOLCANO* DONDE SE POSICIONE EL GEN CORRESPONDERÁ CON UN P-VALOR CONCRETO Y UN VALOR DE FC DETERMINADO. EN ESTE CASO, LOS GENES QUE QUEDEN A LA DERECHA DEL PUNTO 0 DEL EJE DE ABCISAS SERÁN LOS GENES SOBREEXPRESADOS EN EL RD_T2 Y LOS QUE QUEDEN EN EL LADO IZQUIERDO SERÁN LOS SOBREEXPRESADOS EN EL MTX_T3. 136

FIGURA 33. EN EL *VOLCANO PLOT* SE REPRESENTAN TODOS LOS GENES SEGÚN EL P-VALOR AJUSTADO OBTENIDO EN EL AED (EJE DE ORDENADAS) Y EL VALOR DEL LOG₂ FC (EJE DE ABCISAS). CON LAS LÍNEAS PUNTEADAS SE REPRESENTAN DIFERENTES NIVELES DE P-VALOR Y LA ZONA DEL *VOLCANO* DONDE SE POSICIONE EL GEN CORRESPONDERÁ CON UN P-VALOR CONCRETO Y UN VALOR DE FC DETERMINADO. EN ESTE CASO, LOS GENES QUE QUEDEN A LA DERECHA DEL PUNTO 0 DEL EJE DE

ABSCISAS SERÁN LOS GENES SOBREENPRESADOS EN EL PRE_T1 Y LOS QUE QUEDEN EN EL LADO
IZQUIERDO SERÁN LOS SOBREENPRESADOS EN EL MTX_T3. 139

FIGURA 34. A) REPRESENTACIÓN GRÁFICA DE LAS COMPARACIONES REALIZADAS EN BASE A LOS RESULTADOS
OBTENIDOS EN LOS DIFERENTES AED. B) GENES DE COMUNES A LA COMPARACIÓN DE
PRE_T1 vsRD_T2 Y PRE_T1 vsMTX_T3 DISTRIBUIDOS SEGÚN EL LOG2 FC. C) GENES DE COMUNES
A LA COMPARACIÓN DE PRE_T1 vsRD_T2 Y RD_T2 vsMTX_T3 DISTRIBUIDOS SEGÚN EL LOG2 FC.
D) GENES DE COMUNES A LA COMPARACIÓN DE PRE_T1 vsMTX_T3 Y RD_T2 vsMTX_T3
DISTRIBUIDOS SEGÚN EL LOG2 FC. E) GENES DE COMUNES A LA COMPARACIÓN DE PRE_T1 vsRD_T2,
PRE_T1 vsMTX_T3 Y RD_T2 vsMTX_T3 DISTRIBUIDOS SEGÚN EL LOG2 FC. 141



UNIVERSIDAD
DE MÁLAGA

Lista de tablas

TABLA 1. GENES INCLUIDOS EN EL PANEL PAN-CANCER PATHWAYS.	35
TABLA 2. GENES <i>HOUSEKEEPING</i> . LISTADO DE LOS 40 GENES INCLUIDOS EN EL PANEL <i>PAN-CANCER PATHWAYS</i> QUE SE TOMAN PARA LA NORMALIZACIÓN DE LOS DATOS.	37
TABLA 3. RUTAS BIOLÓGICAS QUE ENLOBAN A LOS 730 GENES INCLUIDOS EN EL PANEL.	38
TABLA 4. GENES INCLUIDOS EN EL PANEL DE SECUENCIACIÓN.	45
TABLA 5. CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DE LAS PACIENTES DE LAS COHORTES DONDE LOS MODELOS HAN SIDO DEFINIDOS, <i>COHORTES_DISCOVERY</i>	66
TABLA 6. CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DE LAS PACIENTES INCLUIDAS EN LA COHORTE <i>D_ALLSUBTYPES</i> SEGÚN SI HAN RESPONDIDO O NO AL TRATAMIENTO CON NAC.	67
TABLA 7. CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DE LOS PACIENTES INCLUIDOS PARA LA VALIDACIÓN DE LOS MODELOS PREDICTIVOS, <i>COHORTES_VALIDATION</i>	69
TABLA 8. ASOCIACIONES ENTRE LAS VARIABLES INCLUIDAS EN ESTE ESTUDIO. SOLO SE MUESTRAN AQUELLAS CON SIGNIFICACIÓN ESTADÍSTICA	70
TABLA 9. GENES DE ENTRE LAS PACIENTES QUE OBTIENEN PCR TRAS EL TRATAMIENTO CON NAC Y LAS QUE NO. SOLO SE HAN SELECCIONADO AQUELLOS GENES CON UN P-VALOR AJUSTADO POR DEBAJO DE 0.05. SE INCLUYE UNA BARRA DE INTENSIDAD SEGÚN EL LOG2 FC QUE SE OBTIENE DE CADA GEN. EN ESCALA DE ROJO AQUELLOS GENES CON UN LOG2 FC SUPERIOR A 0, ES DECIR QUE SE EXPRESAN MÁS EN AQUELLOS PACIENTES QUE OBTIENEN PCR Y EN ESCALA DE AZUL AQUELLOS QUE SE EXPRESAN MENOS.	71
TABLA 10. ESTA TABLA DIVIDIDA EN DOS REPRESENTA EL VALOR R DE LA CORRELACIÓN DE <i>PEARSON</i> DE CADA UNO DE LOS PARES DE GENES Y EL P-VALOR DEL T-TEST ASOCIADO A DICHA CORRELACIÓN.	77
TABLA 11. ANÁLISIS DE REGRESIÓN LOGÍSTICA UNIVARIANTE ENTRE LAS VARIABLES CLÍNICO-PATOLÓGICAS INCLUIDAS EN ESTE ESTUDIO Y LOS 10 GENES DE Y LA VARIABLE PCR, ODDS RATIO Y P-VALOR DEL <i>LIKEHOOD RATIO TEST</i> (LRT) DE LAS RELACIONES. SOLO SE MUESTRAN LAS RELACIONES CON UN P-VALOR < 0.05.	80
TABLA 12. RESUMEN DE LOS MODELOS PREDICTIVOS DE LA COHORTE <i>D_ALLSUBTYPES</i> CON LAS VARIABLES SELECCIONADAS (AQUELLAS CON MENOS DEL 5% DE COEFICIENTE IGUAL A 0 EN LA SELECCIÓN CON LASSO) Y EL AUC OBTENIDO EN CADA CASO.	84

TABLA 13. RESUMEN DE LOS MODELOS PREDICTIVOS DE LA COHORTE_D_HER2- CON LAS VARIABLES SELECCIONADAS (AQUELLAS CON MENOS DEL 5% DE COEFICIENTE IGUAL A 0 EN LA SELECCIÓN CON LASSO) Y EL AUC OBTENIDO EN CADA CASO.	88
TABLA 14. RESUMEN DE LOS MODELOS PREDICTIVOS DE LA COHORTE_D_ER+ CON LAS VARIABLES SELECCIONADAS (AQUELLAS CON MENOS DEL 5% DE COEFICIENTE IGUAL A 0 EN LA SELECCIÓN CON LASSO) Y EL AUC OBTENIDO EN CADA CASO.	92
TABLA 15. RESUMEN DE TODOS LOS AUCs OBTENIDOS DE LOS MODELOS PREDICTIVOS TANTO EN LAS COHORTES DONDE HAN SIDO DISEÑADAS COMO EN LA COHORTE DE VALIDACIÓN (INCLUYENDO PAM50 EN LAS COHORTES_ <i>DISCOVERY</i>).	96
TABLA 16. RESULTADOS DE LOS MODELOS DE REGRESIÓN LOGÍSTICA DE COX CON TODAS LAS VARIABLES INCLUIDAS EN ESTE ESTUDIO, SOLO SE MUESTRAN AQUELLAS VARIABLES DONDE EN ALGUNA DE SUS CATEGORÍAS TIENE UNA RELACIÓN ESTADÍSTICAMENTE SIGNIFICATIVA CON EL EVENTO MCE. A. RESULTADOS DEL ANÁLISIS UNIVARIANTE DE COX. B. RESULTADOS DEL ANÁLISIS MULTIVARIANTE DE COX.	99
TABLA 17. CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DE LA DNA-COHORTE.	100
TABLA 18. RECURRENCIA DE LOS GENES MUTADOS EN LAS PACIENTES INCLUIDAS EN ESTE ESTUDIO. ...	101
TABLA 19. CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DE LAS PACIENTES INCLUIDAS EN LA COHORTE PRE_T1 vs RD_T2.	122
TABLA 20. CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DE LAS PACIENTES INCLUIDAS EN LA COHORTE PRE_T1 / RD_T2 / MTX_T3.	123
TABLA 21. GENES DE ENTRE PRE_T1 Y RD_T2.	130
TABLA 22. CLASIFICACIÓN DE LOS 112 GENES DE CON LA HERRAMIENTA DAVID <i>GENE FUNTIONAL CLASSIFICATION TOOL</i>	131
TABLA 23. GENES DE ENTRE RD_T2 Y MTX_T3	135
TABLA 24. GENES DE ENTRE PRE_T1 Y MTX_T3	138

1. Introducción general

En la vida no hay cosas que temer,
sólo hay cosas que comprender.

Marie Curie

1.1. EPIDEMIOLOGÍA DEL CÁNCER DE MAMA

El cáncer de mama (CM) es el tumor más frecuente ente la población, y no solo femenina, sino a nivel mundial. Según los datos proporcionados por GLOBOCAN 2020, la incidencia del CM en la población en general en el año 2020 fue del 11.7%, seguido del cáncer de pulmón (11.4 %) y de colon (10%). En cuanto al número de muertes, es el tipo de tumor que más muertes ocasiona en mujeres (15.5%) (Figura 1, B). En cuanto a la distribución de la incidencia a nivel mundial, según la ratio de edad estandarizada (ASR, *age-standardized rates*), Bélgica tiene el valor de incidencia más alto (113.2 por cada 100.000 mujeres), seguida de Australia, Estados Unidos y Argentina (96, 90.3 y 73.1 por cada 100.000 habitantes, respectivamente) (Figura 1, C). Según las estimaciones en base al crecimiento en la incidencia del CM, la organización mundial de la salud (OMS) estima que en el 2040 habrá un 20.3% más de muertes con respecto al 2020 a causa de esta enfermedad (36).

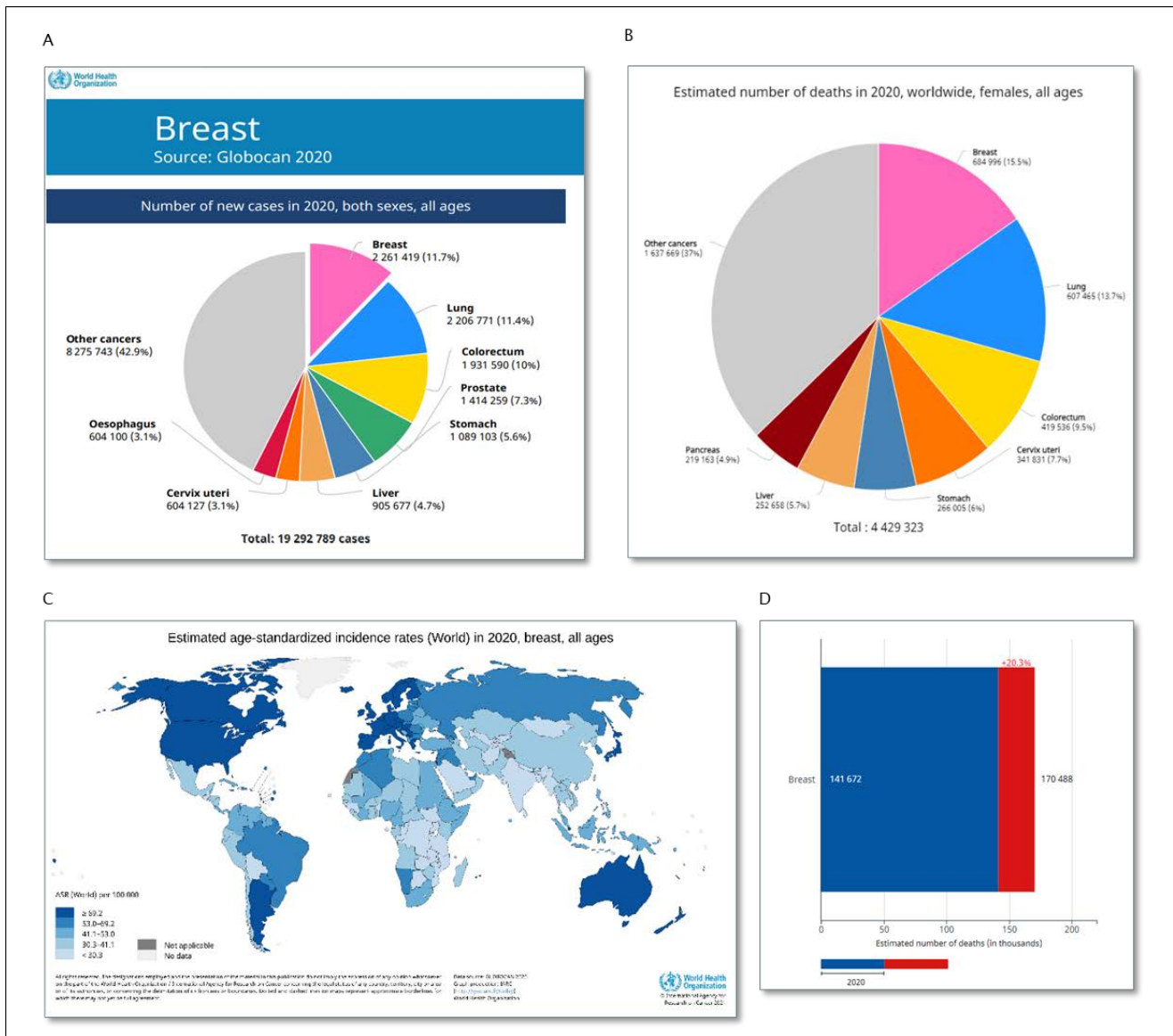


Figura 1. Datos de la fuente *Global Cancer Statistics 2020: GLOBOCAN*.

1.2. HETEROGENEIDAD EN EL CÁNCER DE MAMA: SUBTIPOS Y CLASIFICACIONES

El cáncer de mama es un conjunto de enfermedades heterogéneas que muestra una variación sustancial en su clínica y en sus características moleculares. Existen diversas aproximaciones para capturar dicha heterogeneidad que, además, proporcionan un importante valor pronóstico y predictivo.

1.2.1. Grado y clasificación histológica

El grado histológico evalúa en una escala del 1 al 3 cómo de similar es un tumor con respecto a un epitelio normal: un bajo grado histológico es más similar a un epitelio

normal que un tumor con alto grado (37). La mayoría de los carcinomas de mama se clasifican como carcinomas ductales invasivos (70–80%), mientras que en torno al 25% de los cánceres de mama son caracterizados como “tipos histológicos especiales” que engloban los carcinomas lobulares, tubulares, medulares y metaplásicos (Figura 2). Sin embargo, esta clasificación histológica no refleja la gran heterogeneidad observada en el cáncer de mama; aun así, es un componente esencial en los informes patológicos.

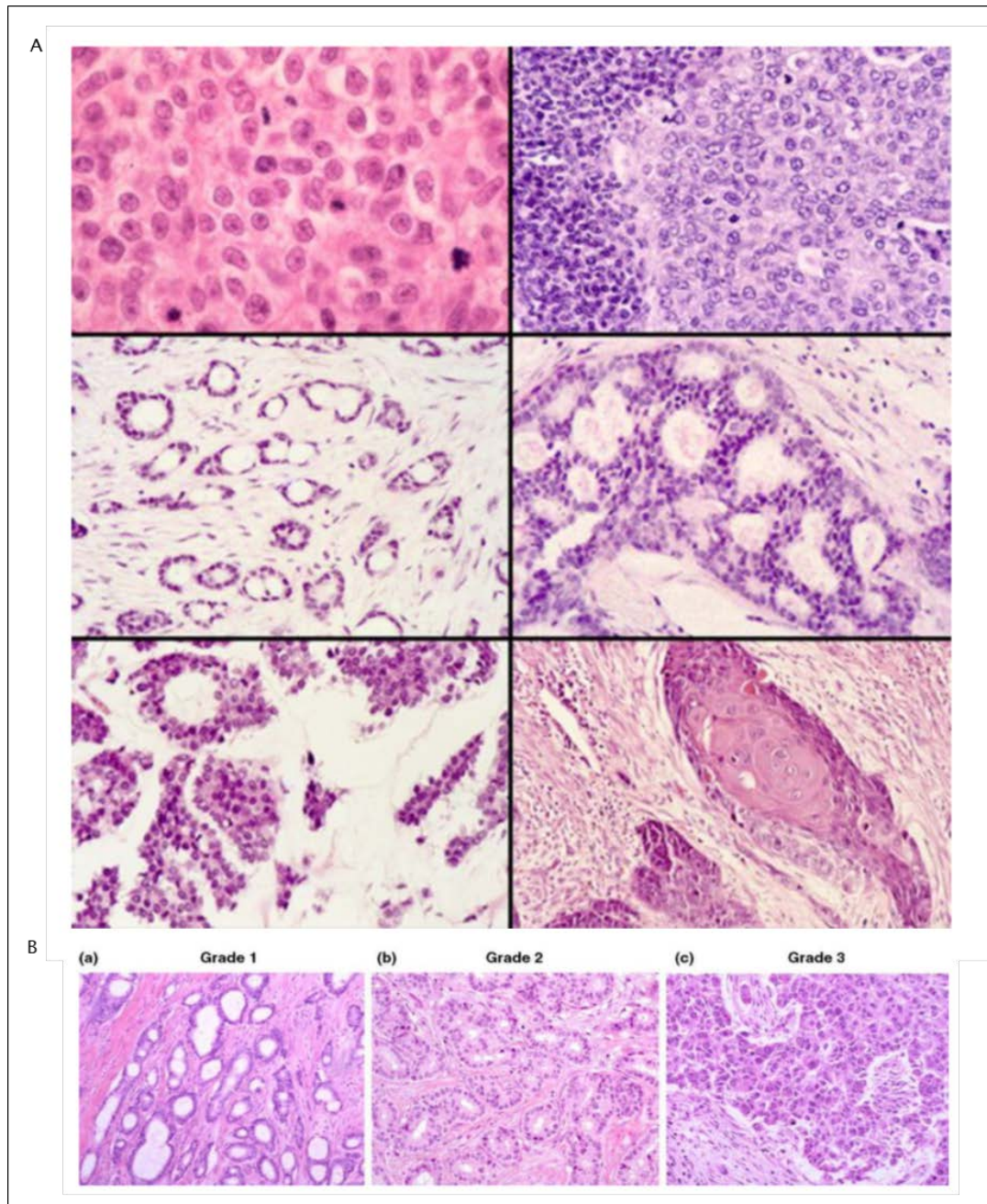


Figura 2. Principales tipos y grados histológicos en el cáncer de mama. A) Tipos histológicos: desde izquierda arriba hasta abajo derecha: cáncer de mama ductal invasivo, medular, tubular, cribiforme, mucinoso y metaplásico. B) Grados histológicos: (a) tumor de mama bien diferenciado (grado 1), (b) tumor de mama moderadamente diferenciado (grado 2) y (c) pobremente diferenciado (grado 3). Figuras obtenidas del Atlas de genética y citogenética en oncología y hematología.

1.2.2. Subtipos Inmunohistoquímicos

Para determinar el pronóstico y las opciones terapéuticas, de manera rutinaria se determina la expresión del receptor de Estrógenos (ER, *estrogen receptor*), el receptor de progesterona (PR, *progesterone receptor*) y la sobreexpresión o amplificación del factor de crecimiento epidérmico humano 2 (HER2, *human epidermal growth factor 2*) que determinan los subtipos inmunohistoquímicos (IHQ) (38). En base a estos biomarcadores, el cáncer de mama se clasifica en tres subtipos: 1) receptor de hormonas positivo o tumores Luminales, que expresan el ER y/o PR (ER+ y/o PR+), 2) tumores HER2-positivo, que sobreexpresan o amplifican el HER2 (HER2+) y 3) tumores triples negativos (TN), que son negativos para los tres biomarcadores (ER-, PR-, HER2-). Un ejemplo de secciones de cáncer de mama inmunomarcadas de los tres subtipos se muestra en la Figura 3. La determinación de estos tres subtipos ha mejorado el manejo de la enfermedad en términos de resultado clínico debido a la administración de terapias dirigidas a los receptores hormonales positivos (terapia hormonal) en el caso de los tumores luminales y al HER2-positivo (terapia anti-HER2) en el caso de los tumores HER2+.

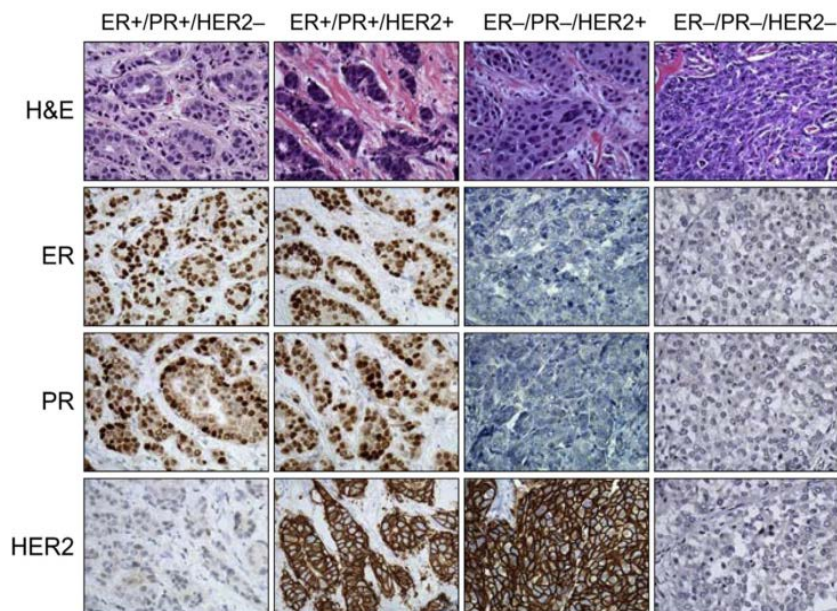


Figura 3. Subtipos de cáncer de mama con tinciones inmunohistoquímicas de los marcadores ER, PR y HER2. La histología del cáncer se ha determinado usando una tinción con hematoxilina y eosina. La imagen procede del artículo de Rivenbark y colaboradores, 2013.

Aunque el uso clínico de estos biomarcadores es muy útil, no predicen respuesta a tratamientos emergentes de nuevas terapias dirigidas ni tienen suficiente valor pronóstico en la toma de decisiones en el tratamiento quimioterapéutico.

1.2.3. Subtipos intrínsecos o moleculares

Los recientes avances en los análisis de expresión génica han llevado al desarrollo de clasificaciones de cáncer de mama basadas en los perfiles de expresión génica. Entre los estudios pioneros se encuentra el de Perou y colaboradores (2000) (39) (40), que propusieron una clasificación que más tarde fue refinada a cuatro categorías con un valor pronóstico predictivo de la respuesta al tratamiento: Luminal A, Luminal B, HER2-enriched y Basal-Like. Estos subtipos son conocidos como “subtipos intrínsecos” debido a que son definidos por sets de genes intrínsecos con una gran variación en la expresión génica entre los diferentes subtipos y no entre muestras del mismo tumor (sus patrones de expresión son característicos de cada tumor y no debido a la toma de la muestra). El subtipo **Luminal A** (del 50 al 60% de todos los cánceres de mama en mujeres) es usualmente positivo para los receptores hormonales (ER+ y/o PR+), tienen baja proliferación (Ki67<14%) y la menor ratio de recaída y por lo tanto mejor pronóstico (40). El subtipo **Luminal B** (del 15 al 20%) es positivo para los receptores hormonales y tiene un perfil de proliferación más alto que el Luminal A pero menos que los subtipos HER2-enriched o Basal-Like, grado histológico alto, peor pronóstico que el Luminal A y una respuesta más baja al tratamiento endocrino que el Luminal A. El subtipo **HER2-enriched** (15 al 20%) se define por la sobreexpresión de HER2 (u otros genes en el amplión 17q12) y morfológicamente está caracterizado por alta proliferación, alto grado histológico y alto grado nuclear. El subtipo **Basal-Like** (del 15 al 20%) es el subtipo más heterogéneo y está compuesto principalmente por tumores que han perdido la expresión de los receptores hormonales y la sobreexpresión/amplificación del HER2 (por ejemplo, los tumores TN en la clasificación de cáncer de mama basada en la inmunohistoquímica). Los tumores Basal-Like expresan altos niveles de citoqueratinas basales, alto perfil de proliferación (Ki67>50%) y presentan una historia natural más agresiva y peor supervivencia que el resto de los subtipos (41) (42).

Basados en esos subtipos intrínsecos, una firma de 50 genes, **PAM50**, fue desarrollada por Parker y colaboradores (43) permitiendo el desarrollo de Prosigna™, un test aprobado por la FDA y con la marca CE que permite identificar el subtipo intrínseco de un tumor y estimar el riesgo de recurrencia de aquellos tumores con receptores de hormonas positivos. Otras pruebas basadas en la expresión génica, como Oncotype DX™ (44), MammaPrint® (45) y Endopredict® (46) predicen la recurrencia, pero no proporcionan información sobre los subtipos intrínsecos del cáncer de mama.

Derivado de este proyecto de tesis, se ha realizado un profundo análisis que ha sido publicado en una revista de impacto, donde se especifican los últimos estudios de la respuesta al tratamiento con quimioterapia Neoadyuvante, los factores predictivos de esta y la caracterización de la enfermedad residual (1). A continuación, se va a detallar la información más importante.

1.3. QUIMIOTERAPIA NEOADYUVANTE EN EL CÁNCER DE MAMA

El tratamiento con quimioterapia neoadyuvante (NAC, *neoadjuvant chemotherapy*), basado en la administración de quimioterapia estándar (Antraciclinas+/-Taxanos más Trastuzumab en aquellos tumores de mama HER2+) antes de la cirugía, es ampliamente usado en el tratamiento del cáncer de mama temprano y localmente avanzado (47). El principal objetivo de la NAC es el incremento de la preservación de la mama (fomenta la reducción del tumor) y la obtención de respuesta completa patológica (pCR, *pathological complete response*), que mejora la supervivencia libre de enfermedad (SLE) y la supervivencia global (SG) del paciente. Además, la implementación de la NAC ha permitido la búsqueda de nuevas dianas terapéuticas y es un excelente modelo in vivo para investigar los mecanismos de resistencia y sensibilidad de las nuevas terapias.

1.3.1. pCR y enfermedad residual

la paciente a largo plazo después de la cirugía. Esta respuesta se determina con un examen patológico del tejido de la mama y de los ganglios linfáticos axilares que son extraídos quirúrgicamente después de la NAC. La respuesta al tratamiento puede ser dicotomizada en la obtención de pCR o la presencia de enfermedad residual (RD, *residual disease*). La pCR se define como la ausencia de células cancerosas invasivas en la mama

o en los ganglios linfáticos (ypT0/Tis ypN0), mientras que la **RD** se define como la presencia de células cancerígenas invasivas en la mama y en los ganglios linfáticos.

Existen diferentes tipos de gradación y clasificación de pCR y de la RD que han sido establecidos dependiendo de la examinación patológica empleada. Las clasificaciones como Miller and Payne (48), y la carga del tumor residual (RCB, *residual cancer burden*) (49) cuantifican la respuesta a la NAC en diferentes grados, donde uno es pCR y el resto corresponde con un espectro de valores que reflejan la extensión de la RD. La clasificación por Miller and Payne, que es la clasificación que vamos a usar en este trabajo como método de medir la respuesta al tratamiento, se relaciona con la SG y con la SLE ($p=0.02$ y $p=0.04$, respectivamente) y es un predictor independiente de la SG (48). La relación entre la respuesta al tratamiento, pCR, y la mejoría en cuanto al resultado clínico a largo plazo de la paciente ha sido establecido (12) y se ha demostrado que es específico de subtipos (3). En el caso de los subtipos HER2+ y TN por IHQ y HER2-enriched y Basal-Like por PAM50 la obtención de pCR se relaciona con un mejor resultado del paciente al largo plazo. En el caso de los subtipos Luminales existe mejoría, pero no es tan notable ya que estos tienen mejor pronóstico que el resto de los subtipos desde el inicio (3). Cuando se obtiene pCR el riesgo de muerte se ve reducido en un 92% en el caso de los tumores HER2+, un 84% en tumores de mama TN y un 71% en los tumores con los receptores hormonales positivos (50) (51). En general, los pacientes que obtienen pCR tras el tratamiento con NAC tienen mucha menos probabilidad de recaer que aquellos que presentan RD (12) (52) (53) (54) (13). El porcentaje general de pCR en el cáncer de mama tras el tratamiento con NAC es del 31%; este porcentaje varía según el subtipo, en aquellos HR+/HER2- es del 12%, en TN del 40% y del 47% en HER2+ (tratado con NAC+Trastuzumab) (13). Teniendo en cuenta el carácter pronóstico de la obtención de pCR, es de vital importancia determinar de antemano qué pacientes van a responder a la NAC, sobre todo en aquellos subtipos donde se ha visto que la ratio de obtención de pCR tras la NAC es bastante baja como en el caso de los tumores ER+, para así poder diseñar la mejor estrategia terapéutica.

1.3.2. Factores predictivos de pCR

La utilidad de la pCR como subrogado del resultado clínico ha llevado a la búsqueda de biomarcadores que predigan pCR, particularmente en el escenario donde la NAC no es efectiva, en el 70% de las pacientes (13). Para que la información sea más clara hemos dividido los factores predictores de pCR en dos; por un lado, aquellos que están relacionados con las características clinicopatológicas de las pacientes, clásicamente usados, y por otro lado aquellos factores determinados por los avances en la biología molecular del CM, detallando, sobre todo, la firma PAM50.

1.3.2.1. Factores clinicopatológicos predictivos de pCR

Los factores predictivos de pCR clinicopatológicos más estudiados son el tamaño tumoral clínico, el grado histológico tumoral, los subtipos IHQ y el porcentaje de Ki67 antes del tratamiento.

El **tamaño tumoral clínico** (cT), se ha demostrado que es un predictor independiente de pCR en diferentes estudios. Tanto en una cohorte de 957 pacientes con CM (2) como en otra de 2046 (52) se demostró que el tamaño tumoral se relacionaba con la obtención de pCR ($p < 0.001$). En aquellos pacientes con un tamaño T1-2 obtienen aproximadamente el doble de pCR que aquellos con un tamaño T3-4 (2). Tamaños de T1-2 fueron predictores independiente de pCR ($p = 0.001$, OR 3.15) (52).

El **grado tumoral histológico** es un predictor de respuesta al tratamiento con NAC (6) (55). El alto grado histológico (G3) fue uno de los mejores predictores de pCR tras el tratamiento con NAC en una cohorte de 560 pacientes con CM ($p < 0.004$) (6). Los tumores con un grado histológico 3 obtienen pCR ($p < 0.0001$) en un 31%, mientras que los grados G1 y G2 raramente consiguen pCR (56).

Debido al mecanismo antimitótico del tratamiento con NAC, el **porcentaje de Ki67** presente en la biopsia diagnóstica antes del tratamiento es un factor predictivo de pCR, sobre todo en aquellos pacientes con los mecanismos proliferativos más activos, como es el caso de tumores de mama HER2+ y TN (57) (8). Aquellos tumores con un porcentaje de Ki67 > 50 obtenían una ratio de pCR mayor que aquellos con Ki67 < 50% (40 vs 19%,

respectivamente, $p=0.0004$) (8). En un estudio realizado en una cohorte TN, aquellos tumores con un alto porcentaje de Ki67 tenían una respuesta de 3.36 veces mayor al tratamiento con NAC que aquellos con un porcentaje de Ki67 bajo (9). Esta variable clinicopatológica no tiene factor predictivo en aquellos tumores donde las vías de proliferación no están activadas, como es el caso en los tumores de CM ER+ (57) (10).

Con respecto al valor predictivo de pCR, es importante enfatizar que la ventaja de supervivencia asociada a la obtención de pCR se limita a los subtipos de CM y que cada uno de ellos tiene una ratio diferente de obtención de esta. Si nos centramos en los **subtipos IHQ**, aquellos tumores con sobreexpresión de HER2 van a obtener la ratio más alta de respuesta al tratamiento ($>47\%$ de pCR), estando estos resultados avalados por numerosos estudios (14) (13) (3) (4). Los tumores TN también obtienen un porcentaje alto de respuesta, mientras que este es el menor en los Luminales (ER+) (4) (5).

1.3.2.2. Factores moleculares predictivos de pCR

1.3.2.2.1. Subtipos intrínsecos determinados por PAM50

Como ya hemos mencionado con anterioridad, PAM50 es una firma basada en la expresión génica de 50 genes que agrupa a los tumores de mama en cuatro subtipos con diferentes características y diferentes ratios de pCR. Por lo tanto, la determinación de los subtipos moleculares por PAM50 es una herramienta para estratificar los pacientes en base de la obtención de pCR específico de cada subtipo.

En los estudios donde se ha analizado el poder predictivo de la pCR en cohortes con una representación de todos los subtipos, los resultados concluían, como era de esperar por lo conocido en los subtipos IHQ, que los subtipos HER2-enriched y Basal-Like presentaban las ratios más altas de pCR ($p<0.001$ y $p<0.001$, respectivamente) (2) (11), y que el subtipo Luminal A predecía las ratios de pCR más bajas comparadas con las del resto de los subtipos (razón de momios (OR, *Odds Ratio*) 0.34 y $p<0.037$) (20).

Los tumores de mama definidos como HER2+ por IHQ han sido los más estudiados desde el punto de vista de la obtención de pCR, además de los subtipos definidos por PAM50, debido sobre todo al estudio de la combinación de NAC con terapias anti-HER2,

como Trastuzumab, Lapatinib y Pertuzumab. Enmarcados dentro de diferentes ensayos clínicos (NOAH, CherLOB, CALGB, PAMELA o BERENICE) los resultados definen un grupo de tumores HER2-enriched dentro de los tumores HER2+ por IHQ, como aquello que van a obtener el ratio más alto de respuesta al tratamiento con NAC cuando son comparados con ese otro grupo de tumores no-HER2-enriched. Dentro de ese grupo no-HER2-enriched, aquellos tumores definidos como Luminal A por PAM50 obtienen la ratio más baja de respuesta al tratamiento (15) (16) (17) (18) (19) (7). El uso del clasificador PAM50 para predecir la respuesta al tratamiento ha demostrado ser más específico que los clásicos subtipos por IHQ, ya que dentro de una cohorte homogénea de HER2+ por IHQ, PAM50 es capaz de distinguir tumores con un comportamiento antagónico con respecto a la respuesta a la NAC (HER2-enriched vs Luminal A).

1.3.2.2.2. Otros predictores moleculares de pCR

Predictores como DLDA-30 (basado en la expresión de 30 genes) o GGI (que combinaba la expresión de 97 genes relacionados con el grado tumoral y otros parámetros clínicos) fueron pioneros en la predicción de pCR tras el tratamiento con NAC con una precisión en la predicción (*accuracy*) de 0.76 y 0.71 respectivamente (58) (59). Desde entonces muchas han sido las mejoras realizadas con el método para refinar la predicción, siendo cada vez más rigurosos, la especificidad y el uso de los avances en la biología molecular.

El estudio de las alteraciones en el número de copias (CNAs, *copy number alterations*), expresión génica y polimorfismos de nucleótido único (SNPs, *single nucleotide polymorphism*) procedente de la información generada de proyectos tan ambiciosos como *The Cancer Genoma Atlas Network* (TCGA), han sido la base de los análisis integrativos que definen 10 clústeres (IntClust 1-10) que redefinen los subtipos de cáncer de mama con diferentes características clínicas y de resultados clínicos (60). Los IntClust 1-2-6-9 podrían definirse como Luminales con mal pronóstico, los 3-4-7-8 como luminales con buen pronóstico, el IntClust 10 como Basal-Like y el 5 como HER2-enriched. El IntClust 2 obtiene las ratios más bajas de pCR (0%) mientras que el IntClust 10 obtiene las ratios más altas de respuesta (37%).

Tal y como ocurre en los subtipos de CM definidos por IHQ, los subtipos moleculares determinados por PAM50 y los IntClust, la información acerca de la predicción de pCR es subrogada del subtipaje previo del cáncer de mama.

Debido a la agresividad que presenta el tumor cuando los subtipos de CM HER2+ y TN no obtienen pCR tras el tratamiento con NAC, se han definido predictores únicos para dichos subtipos. Por ejemplo, en el caso de tumores HER2+ tratados con NAC-Trastuzumab, se ha definido un predictor de respuesta al tratamiento llamado *Trastuzumab Risk model* (TRAR) basado en la expresión de 41 genes donde el poder de predicción de pCR o área bajo la curva (AUC, *area under the curve*) era de 0.73. Este poder de predicción variaba según si el tratamiento con Trastuzumab era sustituido por Lapatinib (AUC 0.76). Cuando la información de TRAR es combinada con las variables clinicopatológicas, el valor predictivo aumentaba hasta superar al clasificador PAM50 (AUC 0.78 vs 0.74), siendo un ejemplo de la sinergia que debe haber entre la clínica y los estudios moleculares (61).

En el caso de tumores de mama TN, Lehmann y colaboradores, definió seis entidades moleculares dentro de este subtipo, con diferentes características moleculares y respuesta al tratamiento: Basal-Like 1 (BL1), Basal-Like 2 (BL2), Inmunomodulatory (IM), Mesenchymal (M), Mesenchymal stem-like (MSL) y Luminal androgen receptor (LAR) (62) (63) (64). El subtipo BL1 tiene el perfil más proliferativo con un porcentaje de Ki67 > 50%, y obtiene la ratio mayor de pCR, sobre todo en pacientes tratados con NAC y sales de platino (80%). El subtipo LAR, que corresponde a tumores HER2-enriched y Luminal A por el clasificador PAM50, es el subtipo con menor ratio de respuesta al tratamiento con NAC (14.3%) (64).

1.4. ENFERMEDAD RESIDUAL EN CÁNCER DE MAMA TRAS EL TRATAMIENTO NEOADYUVANTE

Como hemos mencionado previamente, la presencia de células tumorales después del tratamiento Neoadyuvante es un factor pronóstico fundamental. Muchos estudios revelan que el grado de reducción del tumor después de la NAC está asociado con un incremento de la SLE y de la SG (12) (3) (53) (13). A continuación, vamos a describir los

diferentes biomarcadores de la RD que son usados para determinar el resultado a largo plazo de los pacientes tras el tratamiento con NAC y de la cirugía.

1.4.1. Determinación del Ki67 después del tratamiento Neoadyuvante

En 2009, Jones y colaboradores fueron pioneros en determinar el resultado del poder predictivo del marcador de proliferación Ki67 tras el tratamiento con NAC. En una cohorte de 130 pacientes con muestras pareadas entre tumor primario y RD, llegaron a la conclusión de que la reducción de Ki67 después del tratamiento con NAC era un fuerte predictor del resultado clínico en pacientes que no consiguen pCR ($p < 0.001$; HR 1.6) (65). Esta reducción en Ki67 después del tratamiento con NAC fue correlacionada con una mejor DFS y OS, particularmente en el subtipo Luminal B. Además, el análisis multivariante mostró que la falta de reducción en Ki67 durante el tratamiento con NAC incrementaba significativamente el *Hazard ratio* (HR) de la recurrencia y de la muerte (3.39 y 7.03, respectivamente) (10).

En el ensayo clínico GeparTrio, donde se evaluó el valor pronóstico del índice de Ki67 tras el tratamiento con NAC, 1151 pacientes con cáncer de mama (484 con pCR y 667 con RD), fueron divididas en tres grupos según su % de Ki67; el grupo de Ki67 bajo (0–15%), el grupo de Ki67 medio (15.1–35%) y el grupo de Ki67 alto ($> 35.1\%$). El resultado del tratamiento a largo plazo del grupo con Ki67 bajo fue equiparable al grupo de pacientes que obtenían pCR ($p < 0.211$ para la SLE y $p < 0.779$ para la SG). El grupo de pacientes con Ki67 alto tuvo más riesgo de recaída de la enfermedad ($p < 0.0001$) y muerte ($p < 0.0001$) que los grupos de pacientes con Ki67 bajo y medio (57) (66).

1.4.2. Índice de carga residual de cáncer

Como hemos mencionado con anterioridad, el no conseguir pCR tras el tratamiento con NAC está relacionado con un mal pronóstico a largo plazo (12) (3) (53) (13). Por esta razón, la RD ha sido exhaustivamente caracterizada para definir una estratificación de pacientes acorde al resultado a largo plazo del tratamiento con NAC. El RCB es un sistema que cuantifica la enfermedad residual basada en seis parámetros como son: la celularidad, el porcentaje de carcinoma in situ, el número de nodos metastásicos, el diámetro de la lesión metastásica más grande y el tamaño del tumor residual que queda

tras el tratamiento. Se han determinado 4 grupos diferentes según el resultado de los pacientes a largo plazo: RCB-0 (pCR), RCB-I (mínima RD), RCB-II (moderada RD) y RCB-III (RD extensa). Symmans et al. (2007) aplicaron este sistema a una cohorte de 382 pacientes con diferentes tratamientos y demostraron que el RCB era un predictor independiente de Supervivencia Libre de Recaída a Distancia (SLRD), siendo además reproducible y costo-efectiva. Los índices RCB-0 y RCB-I tenían similar SLRD (5.4% y 2.4% respectivamente), mientras que el índice RCB-III se asociaba al mayor porcentaje de SLRD (53.63%) (49). Esto a priori no parece tener mucho sentido ya que obtener pCR está relacionado con una buena respuesta a largo plazo y esta es equiparable con el índice RCB-0, pero cuando caracterizamos grupos según el subtipo de cáncer de mama, vemos que en aquellos tumores HR+/HER2- con RCB-III son en su mayoría Luminal A definidos por PAM50 ($p < 0.01$) (67), que aunque no respondan tan bien al tratamiento con NAC, como el HER2-enriched o el Basal-Like, sí que han sido relacionados con un buen resultado del paciente a largo plazo. Consistente con el hecho de que aquellos tumores ER+ y PR+ tienen menos probabilidad de conseguir pCR, en un estudio reciente se observó que los genes codificantes de dichos receptores (*ESR1* para ER y *PGR* para PR) estaban sobreexpresados en pacientes con RCB-II y RCB-III y no lo estaba en aquellos pacientes con pCR o RCB-0/I ($p = 0.00053$ y 0.0061 , respectivamente). *ESR1* estaba más expresado en RCB-III con respecto a RCB-0/I y RCB-II ($p = 0.041$). Por otra parte, el nivel de expresión de *MKI67* fue más bajo en RCB-III que en RCB-0/I y RCB-II (0.0029). Finalmente, en el subtipo TN, y probablemente relacionado con su desfavorable pronóstico, el índice RCB-III fue más asociado con un alto estadio clínico y alto ratio de ganglios linfáticos afectados que los índices RCB-0/I y RCB-II ($p < 0.01$) (67).

Mientras que el índice RCB podría ser un espejo de los subtipos moleculares de cáncer de mama definido por PAM50 en términos de respuesta al tratamiento con NAC, otras características del tumor están asociadas, de manera independiente, al resultado del tratamiento con NAC y por tanto deberían ser consideradas transversalmente al índice RCB cuando caracterizamos tumores tras el tratamiento. Por ejemplo, en una cohorte de 109 pacientes con tumores TN con enfermedad residual, el análisis multivariante indicaba que tanto los niveles en el estroma de linfocitos infiltrantes en tumor (TILs,

Tumor infiltrating lymphocytes) CD4+ como el índice RCB tenían valor pronóstico independiente para el intervalo libre de recurrencia a distancia (ILRD) (TILs CD4+: HR 2.88, $p < 0.007$ y el índice RCB: HR 12.04, $p < 0.0001$) (68).

1.4.3. Carga residual proliferativa del cáncer

La carga residual proliferativa del cáncer o RPCB (*Residual Proliferative Cancer Burden*) aglutina tanto los eventos de Ki67 como los del RCB tras la cirugía. La mejoría en cuanto al valor pronóstico del RPCB con respecto a los valores individuales de Ki67 y RPCB es notable. El RPCB agrupa con más eficacia a los grupos de pacientes con diferentes ratios de SG y SLE (57) (69). Cabe destacar que el valor pronóstico difiere entre tumores de mama ER+ y ER-, de manera que todas las variables fueron relacionadas con el resultado clínico a largo plazo en el grupo de tumores ER- ($p < 0.001$), pero, en tumores ER+, donde Ki67 y RCB carecían de valor pronóstico, el RPCB sí lo tenía ($\chi^2 = 13.3$, $p < 0.003$) (69).

1.4.4. Dinamismo molecular de la RD después del tratamiento con NAC

En el escenario post tratamiento, es particularmente importante identificar las alteraciones moleculares inducidas por el tratamiento para poder así relacionarlas con los mecanismos de quimio-resistencia. La hipótesis predominante ha sido que ocurre una selección de las clonas celulares cancerosas que son resistentes al tratamiento, pero es importante caracterizar el dinamismo molecular del tumor durante el tratamiento para determinar y definir la resistencia inducida por NAC más allá de los predictores basales de pCR. Esto permitiría el diseño y el estudio de tratamientos más específicos y, conseguir así mejores resultados clínicos (57).

Hasta un 30% de las pacientes tratadas con NAC, incluyendo a las pacientes HER2+ tratadas con NAC-Trastuzumab, presentan cambios en el estatus de los biomarcadores ER y HER2 durante el tratamiento (57) (70). Además, esas variaciones son dependientes del tratamiento. Por ejemplo, el 43% de las pacientes HER2+ pierden la positividad de dicho marcador tras el tratamiento con NAC-Trastuzumab en comparación con el 9,5% de las pacientes tratadas sólo con NAC (57). Dicha pérdida del estatus positivo de HER2 después del tratamiento diana antiHER2 como es el caso de Trastuzumab, podría

deberse a la erradicación de manera específica de las clonas HER2+, permitiendo la identificación de clonas HER2- preexistentes desde el inicio atribuidas a la alta heterogeneidad intratumoral (57). La pérdida de la positividad del biomarcador HER2 tras la NAC ha sido relacionada con un una peor SLE y peor supervivencia libre de recaída (SLR) ($p=0.041$) comparada con aquellos pacientes que mantienen el estatus HER2+ (57) (71) (70).

En cuanto a los receptores hormonales ER y PR, la pérdida ocurre más que la ganancia tras el tratamiento con NAC (72). Esta ganancia está relacionada con una mejoría en cuanto al resultado clínico a largo plazo cuando se compara con la pérdida o la ausencia de variación, debido, sobre todo, a la adquisición de una nueva diana terapéutica ($p=0.039$ para SLE; $p=0.045$ para SG). Estos cambios a nivel de biomarcadores clásicos en el cáncer de mama, también se han visto en los subtipos determinados por PAM50; por ejemplo, los cambios más significativos han sido de Luminal B a Luminal A, debido en gran medida a la eliminación de aquellas clonas más proliferativas (73) (74).

En base a esto, una nueva determinación de los biomarcadores que son negativos en el tumor primario debería ser priorizada y, además, motiva a determinar a la RD como una nueva entidad tumoral más que como los vestigios del tumor primario.

1.4.5. Perfil transcriptómico y genómico de la RD tras la NAC

Las modificaciones encontradas en una cohorte de 56 pacientes donde se analizaron las diferencias encontradas entre el tumor primario y la enfermedad residual a nivel transcriptómico, evidenciaba una diferencia entre los diferentes subtipos de cáncer de mama definidos en el diagnóstico. En la población de CM en general (64% Luminales y 36% TN) se observó una disminución de los genes de proliferación después de la NAC, en genes como MKI67, TOP2A y AURKA. También aquellos genes marcadores de células madre como *CD44* y *STAT3* y los genes *BCL2*, *PLAU*, *STAT1* estaban desregulados Como genes más expresados tras la NAC se encontraban *SNAI1*, *SNAI2*, *SOX9* y *TWIST*, que están relacionados con la Transición Epitelio-Mesénquima (EMT, *epithelial-mesenchymal transition*), los genes apoptóticos *PAWR* y *BCL2*, los genes relacionados

con la vía RAS-ERK como es el caso de *DUSP4*, los genes *ACACB*, *LDHB* y *ALDH1A1* relacionados con el metabolismo y los genes estromales *DECORIN* y *SPARC*.

1.4.5.1. Alteraciones moleculares en la RD de tumores TN

Debido a la agresividad que presenta el subtipo TN cuando el tratamiento con NAC no es efectivo, la mayoría de los estudios se han centrado en este subtipo. Cuando los tumores de mama TN no consiguen pCR y por tanto presentan RD tras el tratamiento con NAC, aumenta el riesgo de recurrencia metastásica y el resultado clínico a largo plazo del paciente es peor con respecto al resto de subtipos (75). La mayoría de las alteraciones que caracterizan a la RD de estos tumores se encuentran en *TP53* (89%), pérdidas de copias génicas en *BRCA2* (68%), seguido de amplificaciones génicas en *MCL1* (54%) y *MYC* (35%) y mutaciones en *PIK3CA* (11%) (75) (76) (77). Amplificaciones génicas en *GATA3* y pérdidas génicas en *FLT3* y *RB1* también han sido encontradas en la RD (76). En los estudios donde se han comparado las alteraciones moleculares encontradas entre el tumor primario con respecto a la RD, se determinó que las amplificaciones en *MCL1* son más frecuentes en la RD de tumores TN que en los tumores primarios Basal-Like: 54% en RD versus 19% en el tumor primario antes del tratamiento ($p=0.0006$), lo que podría indicar que *MCL1* juega un papel importante en la resistencia a la NAC (75). Por otra parte, se han encontrado más cambios en los CNAs que con respecto a los cambios de un único nucleótido entre el tumor primario y la RD de tumores TN tratados con NAC. La NAC modula un patrón de CNAs con una adquisición caótica de ganancias y pérdidas, incluyendo un aumento de la amplificación de genes conocidos como son *MCL1*, *MYC*, *KRAS*, *BCL9*, *EGFR*, *CDKN2A*, *BIRC2* y *BIRC3* en la enfermedad residual con respecto al tumor primario. En tumores de mama TN la alteración genética de *TP53* se ha demostrado que tiene un impacto importante en el futuro del paciente, de manera que una pérdida de *TP53* se encuentra asociado a una significativa disminución en SLE y SG ($p<0.02$ y $p<0.03$, respectivamente) (24). Determinadas mutaciones en *TP53* y la activación de la vía de señalización de *MYC* tienden a ser concomitantes en dicho subtipo ($p<0.05$) y ambas alteraciones se asocian con una disminución en la SLE y SG ($p=0.007$ y $p=0.007$, respectivamente). También cabe destacar que la ganancia de copias en la región próxima al gen *SMAD4* está asociada a una peor SLE en este tipo de tumores (24).

En cuanto a los cambios encontrados en la expresión génica después del tratamiento con NAC en los subtipos de mama TN, tras el análisis de 450 transcritos, Hancock y colaboradores (2019), determinaron que se podían diferenciar dos tipos de tumores en la RD, unos tumores residuales donde predominaban la activación de genes involucrados en la diferenciación, alto % de Ki67 y bajo perfil de la actividad de las vías TGF β y MEK, mientras que otros tumores residuales presentaban un enriquecimiento en genes relacionados con vías de invasión y una escasa diferenciación. Consistente con la idea de la supervivencia de clonas con características de células madre o *stem-like* tras el tratamiento con NAC (24), este grupo se caracterizaría por un aumento en la activación de las vías MEK y TGF β junto con la delección del gen *DUSP4* (75) (78) (79) (80). De hecho, algunos estudios enfocados en el subtipo TN han coincidido en la reducción de la expresión del gen *DUSP4* en la enfermedad residual y además la han relacionado con peor SLR ($p=0.0004$) (78). Con las mismas conclusiones, un estudio que evaluaba la expresión génica en RD de tumores de mama TN determinó la posibilidad de definir dos tipos de tumores, aquellos con mayor expresión de genes relacionados con la luminalidad como *AR* o *GATA3*, que presentaban un aumento importante en la SLR que aquellos tumores con mayor expresión de genes marcadores de células madre como es el caso de *KRT16*, *WNT11* o *MMP28* (77% versus 25%). Por último, otro estudio con el objetivo de definir una firma que defina el resultado clínico del paciente en base al perfil transcripcional de la RD, determinó que la expresión de los genes *CCL5*, *DDIT4* y *POLR1C* poseían valor pronóstico para SLRD en este subtipo de CM ($p=0.002$) (81). En el estudio individual de los genes, la sobreexpresión del gen *CCL5* fue relacionada con buen pronóstico ($p=0.0012$) mientras que la sobreexpresión de *DDIT4* fue asociada al mal pronóstico ($p=0.00034$) en relación a la SLR (82).

1.5. TUMOR PRIMARIO, ENFERMEDAD RESIDUAL Y METÁSTASIS

Más allá de la RD, este campo carece en gran medida de factores deterministas de metástasis después del tratamiento con NAC y su interrelación con perfiles específicos del tumor primario y la RD. Es así una preocupación importante, más allá de la detección

y predicción de la RD, el desarrollo de la enfermedad metastásica, ya que es la responsable del 90% de las muertes relacionadas con el cáncer de mama (25).

Actualmente, algunos estudios han establecido algunos biomarcadores candidatos de la metástasis en cáncer de mama; sin embargo, no hay predictores de metástasis que podamos identificar en la RD tras el tratamiento con NAC. Es por ello por lo que los estudios longitudinales con cohortes homogéneas controladas con la obtención o no de pCR son la clave para la identificación del impacto de la NAC en el desarrollo de metástasis a distancia y le evolución tumoral hacia fenotipos más deletéreos de esta enfermedad.

2. Hipótesis y objetivos

La ciencia, para mí, explica parte de la vida. Hasta donde
llega, se basa en hechos, experiencias y experimentos.

Rosalind Franklin

2.1. HIPÓTESIS

Existen determinados perfiles moleculares que predicen de forma diferencial la respuesta al tratamiento con NAC en el cáncer de mama primario en los diferentes subtipos de cáncer de mama sin ser necesariamente marcadores subrogados de los mismos.

Los cambios en el perfil molecular observado en el tumor primario con respecto a la enfermedad residual después del tratamiento con NAC permiten identificar alteraciones moleculares potencialmente explotables desde el punto de vista terapéutico.

Por otra parte, el perfil molecular en el tumor residual predice de forma consistente el perfil molecular del tumor metastásico.

2.2. OBJETIVOS

Objetivo 1: Determinar una firma predictiva de pCR que identifique antes del tratamiento aquellas pacientes que no van a responder a la quimioterapia Neoadyuvante con mayor especificidad e independencia del subtipo que las actuales herramientas de predicción.

Objetivo 2: Identificar las alteraciones moleculares comunes y específicas del tumor primario y el tumor residual tras el tratamiento con NAC para entender la evolución tumoral e identificar marcadores específicos de resistencia.

Objetivo 3: Establecer el espectro de variantes moleculares asociadas al primer nicho metastásico para establecer los marcadores específicos de metástasis y aquellos predictores de la misma en la enfermedad residual.

3. Material y métodos generales

3.1. PACIENTES Y TRATAMIENTOS

Con el objetivo de reclutar el mayor número de pacientes para responder a las preguntas del presente estudio, se inició una colaboración con el Hospital Clínico de Valencia, donde se reclutaron seis pacientes para responder al objetivo 2.

El tratamiento con NAC se basa en la administración de quimioterapia estándar (Antraciclinas+/-Taxanos más Trastuzumab en aquellas pacientes con sobreexpresión del HER2). Se han seleccionado pacientes en 3 tiempos diferentes: Pre_T1 serán aquellas muestras del tejido tumoral antes del tratamiento con NAC (procedente de la biopsia diagnóstica). RD_T2 designa a las muestras del tejido tumoral tras el tratamiento y posterior cirugía, es decir, de la enfermedad residual. Y Mtx_T3 a aquellas muestras de tejido tumoral procedente del primer nicho metastásico. En el caso de que la paciente haya conseguido pCR tras el tratamiento con NAC, no hay muestras RD_T2 y Mtx_T3.

Han sido reclutadas para el presente trabajo un total de 111 pacientes diagnosticadas con cáncer de mama, de las cuales todas han sido tratadas con NAC y, en aquellas pacientes con sobreexpresión de HER2, NAC más Trastuzumab (anti-HER2). En cuanto a los análisis de expresión, contamos con tres agrupaciones de nuestra cohorte de pacientes según el objetivo. En el contexto del objetivo 1, de estas 111 pacientes se seleccionaron 84 pacientes en Pre_T1 con la información clínica completa. De estas 84 pacientes Pre_T1, 28 habían conseguido pCR y 56 presentaban enfermedad residual tras el tratamiento. Para ser más robustos en nuestros resultados, en esta parte del trabajo hemos usado un set de datos de otras cohortes publicadas para validar nuestros datos. El número de pacientes para la validación varía según los marcadores de cáncer de mama de que se disponga. Esta información viene detallada en la figura 4 y desarrollada en el capítulo I.

En el capítulo II, de las 111 pacientes, se seleccionaron 70 pacientes que no habían conseguido pCR después de la NAC+/-Trastuzumab, con muestras pareadas de antes y después del tratamiento (Pre_T1 y RD_T2). En el objetivo 3, de estas 70 pacientes seleccionadas para el objetivo 2, conseguimos muestra de tejido tumoral del primer nicho metastásico de aquellas pacientes que en el momento del experimento tenían recaída a distancia (22 pacientes de las se disponía de muestra en Pre_T1, RD_T2 y

Mtx_T3). No disponemos de la información adicional sobre si el resto de las pacientes incluidas en este objetivo presentaron recaída a distancia hasta el momento de la escritura de la presente tesis. En la figura 4 se describe la distribución de las pacientes.

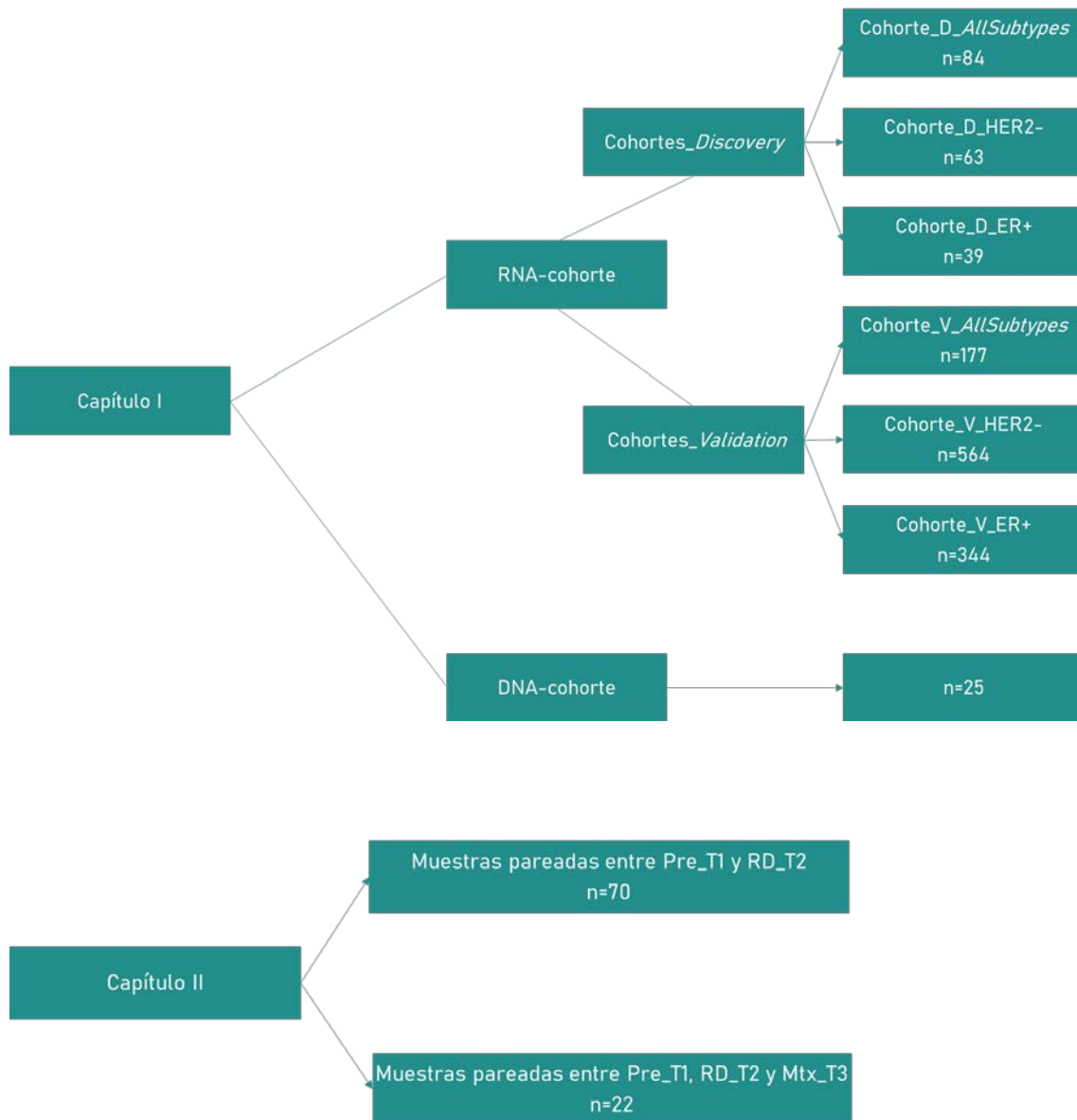


Figura 4. Distribución de las pacientes en los capítulos I y II.

3.2. EVALUACIÓN HISTOLÓGICA, pCR Y SUBTIPOS IHQ DE MAMA.

El nodo del Biobanco Andaluz en Málaga que gestiona las muestras del Hospital-IBIMA nos proporcionó la mayoría de las muestras tumorales que han sido usadas en este proyecto. Para completar el set de pacientes y poder llevar a cabo el objetivo 3, el Hospital Clínico de Valencia colaboró con nuestro grupo.

Las muestras han sido recuperadas del remanente de la biopsia diagnóstica (en el caso de Pre_T1 y de Mtx_T3) o durante la cirugía (para el RD_T2). Son muestras incluidas en bloques de parafina (FFPE, formalin-fixed paraffin-embedded), para así mantener la morfología celular y preservar el tejido por más tiempo.

Para delimitar el área del tumor, y descartar el tejido sano subyacente al tumoral, se realizó una macro-disección manual. De cada bloque de parafina se hizo un corte de 3 micras para, posteriormente, teñirla con una tinción de Hematoxilina y Eosina (H&E), identificando así las estructuras básicas como el núcleo celular con tonos azulados (Hematoxilina) y las estructuras ácidas como el citoplasma con tonos rosados (Eosina). Esto le permite al especialista de Anatomía patológica delimitar la zona tumoral y calcular la celularidad de la muestra, que, para los experimentos de biología molecular como es el caso del presente trabajo, tenía que ser mayor o igual al 30%. Todas aquellas muestras con una celularidad por debajo del 30% fueron descartadas.

El mismo procedimiento fue llevado a cabo para **determinar la respuesta al tratamiento o pCR**. Con una tinción de H&E el patólogo determinó un valor dentro de la clasificación de Miller and Payne, donde el 5 equivale a ausencia de células tumorales y del 4-1 equivale a diferentes grados de presencia de células tumorales. Es decir, todas aquellas pacientes con un valor 5 en la clasificación de Miller and Payne tanto en la muestra de la mama como en la muestra del ganglio linfático de la axila (ypT0/Tis, N0), se consideran buenas respondedoras al tratamiento con NAC y presentan pCR. Aquellas pacientes con valores entre 4 y 1 en dicha clasificación (en la muestra de la mama o en la axila) se consideran pacientes con tumores resistentes al tratamiento con NAC o, dicho de otra manera, pacientes con enfermedad residual.

Para la **identificación de las proteínas** tales como el ER, el PR, las citoqueratinas 5/6 (CK5/6), el receptor del factor de crecimiento epidérmico (EGFR), el Ki-67 y el HER2, se

realizaron tinciones inmunohistoquímicas en secciones del bloque de parafina de 5 µm. Se usó una estación de tinción automatizada (Autostainer Plus, Dako), utilizando el EnVision FLEX System (Dako). Como cromógeno, usamos diamobenzidina y para reemplazar el anticuerpo (Ac) primario se usó un suero no-inmune como un control negativo, y un control positivo específico para cada anticuerpo.

Los anticuerpos que se usaron para la determinación de las proteínas específicas fueron: ER (Ac monoclonal de conejo, clona SP1, Master diagnostica); PR (Ac monoclonal de conejo, clona Y85, Master diagnostica); CK5/6 (Ac monoclonal de ratón, clona D5/16B4, Master Diagnostica); EGFR (Ac monoclonal de conejo, clona EP38Y, Master Diagnostica); Ki-67 (Ac monoclonal de conejo, clona SP6, Master Diagnostica); y HER2 (Kit Herceptest, Dako). La tinción fue evaluada por el grupo de anatomía patológica. Para ER y PR el resultado positivo fue toda aquella tinción nuclear por encima del 1% (83) y para EGFR y CK5/6 cualquier tinción membranosa o citoplasmática. El Ki67 fue determinado cuantificando el porcentaje de células neoplásicas núcleo-positivas. La sobreexpresión de HER2 fue evaluada de acuerdo a las guías internacionales (84) y categorizadas como negativas si el resultado de la inmunohistoquímica era 0 o +1; cuando el resultado era +2 se determinó la amplificación del gen por hibridación *in situ* fluorescente (FISH) o por hibridación *in situ* cromogénica (CISH).

Por lo tanto, los tumores fueron **clasificados inmunohistológicamente** como **Luminal A** (ER-positivo y/o PR-positivo, HER2-negativo y KI67<14%), **Luminal B** (ER-positivo y/o PR-positivo, HER2-negativo y KI67>14%), **HER2-positivo** (HER2-positivo, independientemente del valor de ER y PR) y **Triple negativo** (ER-negativo/PR-negativo/HER2-negativo, CK5/6-positivo y/o EGFR-positivo).

Otras variables clínicas que fueron recogidas como parte del estudio son: estado menopáusico en el momento del diagnóstico, el tamaño tumoral clínico, el grado histológico del tumor primario, la afectación ganglionar axilar antes del tratamiento, tipo de NAC, fecha de cirugía, tipo de cirugía, tamaño tumoral final, número de ganglios linfáticos afectados (en el caso de que los haya), si ha recibido hormonoterapia después del tratamiento con NAC, recidiva y órgano de recidiva. Todas estas variables fueron recopiladas de la historia clínica de las pacientes que están recogidas en Galén (85), un

sistema de información para la gestión y coordinación de la información clínica de nuestro hospital.

En cada uno de los capítulos de la presente tesis doctoral se proporcionará en la sección de metodología específica una descripción detallada de las variables clínicas de las pacientes que han sido incluidas.

3.3. ESTUDIO DE LA EXPRESIÓN GÉNICA

3.3.1. Aislamiento del mRNA desde el tejido incluido en parafina.

Para la extracción del RNA mensajero se llevó a cabo una macro-disección del tejido tumoral previamente delimitada por el patólogo. La purificación del RNA se llevó a cabo con 3 secciones de 10 μm del tejido FFPE con el kit *High Pure FFPE RNA isolation kit* (Roche Life Science) según la recomendación de Nanostring para el uso de su tecnología nCounter® en el estudio de la expresión génica. Para determinar la pureza y la concentración del RNA mensajero aislado se usó un espectrofotómetro de bajo volumen (Nanodrop, Thermo Scientific). La ratio entre 260nm/280nm (A260/280), que determina la pureza de los ácidos nucleicos con respecto a las proteínas, se han dado por válidos valores próximos a 2. Para identificar una posible contaminación por sales o alcoholes derivados del proceso de aislamiento también se tiene en cuenta la ratio 260nm/230nm (A260/230), las muestras debían estar en torno a 2 para ser incluidas.

3.3.2. Determinación de la expresión génica con la tecnología nCounter Dx Analysis System (Nanostring technologies)

Este sistema mide la expresión génica de los genes a estudiar de forma directa en base al código de color que poseen las sondas a las que se unen cada uno de ellos. Estas sondas pueden ser contadas individualmente dando una alta precisión y sensibilidad y evitando así los posibles sesgos de los métodos basados en la amplificación como el caso de la cuantificación de la expresión génica por qPCR o Microarray (Figura 5). Con esta tecnología se ha medido la expresión génica de 730 genes asociados a cáncer con el panel comercial *PanCancer Pathways*, así como se han determinado los subtipos

moleculares de mama con el panel comercial PAM50, ambas metodologías explicadas a continuación.

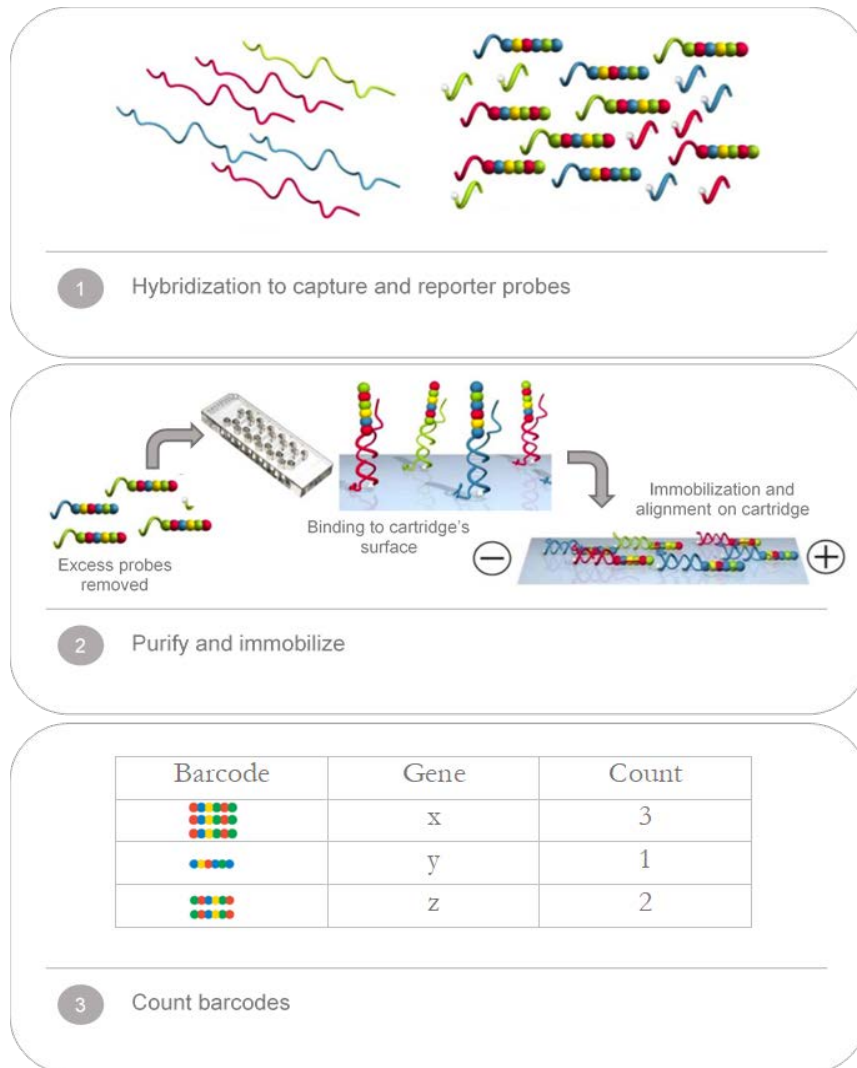


Figura 5. Flujo de trabajo del sistema nCounter®. 1) Cada RNA mensajero hibrida con una sonda reporter (que lleva la señal) y con una sonda de captura (que se encuentra inmovilizada en el cartucho). 2) Las muestras son cargadas en la estación de preparación o PrepStation, donde el exceso de sondas se retira y el complejo RNA mensajero-sonda reporter-sonda de captura es inmovilizada y alineada en la superficie del cartucho. 3) Con el analizador digital o Digital analyzer, se recogen los datos de nuestras muestras; los códigos de color de la superficie del cartucho son contadas y asignadas a cada diana. La imagen ha sido tomada de la web <https://www.nanostring.com/scientific-content/technology-overview/ncounter-technology>.

3.3.2.1. Subtipos moleculares de mama determinados por PAM50

Para determinar los subtipos moleculares de mama **Luminal A**, **Luminal B**, **HER2-enriched** y **Basal-Like**, se usó la tecnología nCounter Dx Analysis System de Nanostring (explicada en el apartado 3.2.2) con el panel “PAM50 signatures”. Este panel determina la expresión de 50 genes (Figura 6) que, en base a un algoritmo, asigna a cada una de las muestras un subtipo. Los criterios propios de dicho panel exigían que las muestras tuvieran una concentración mínima (>12.5 ng/uL) así como unos mínimos de calidad (ratio de A260/280 entre 1.7 y 2.3); todas aquellas muestras que no cumplieran con estos requisitos fueron descartadas.

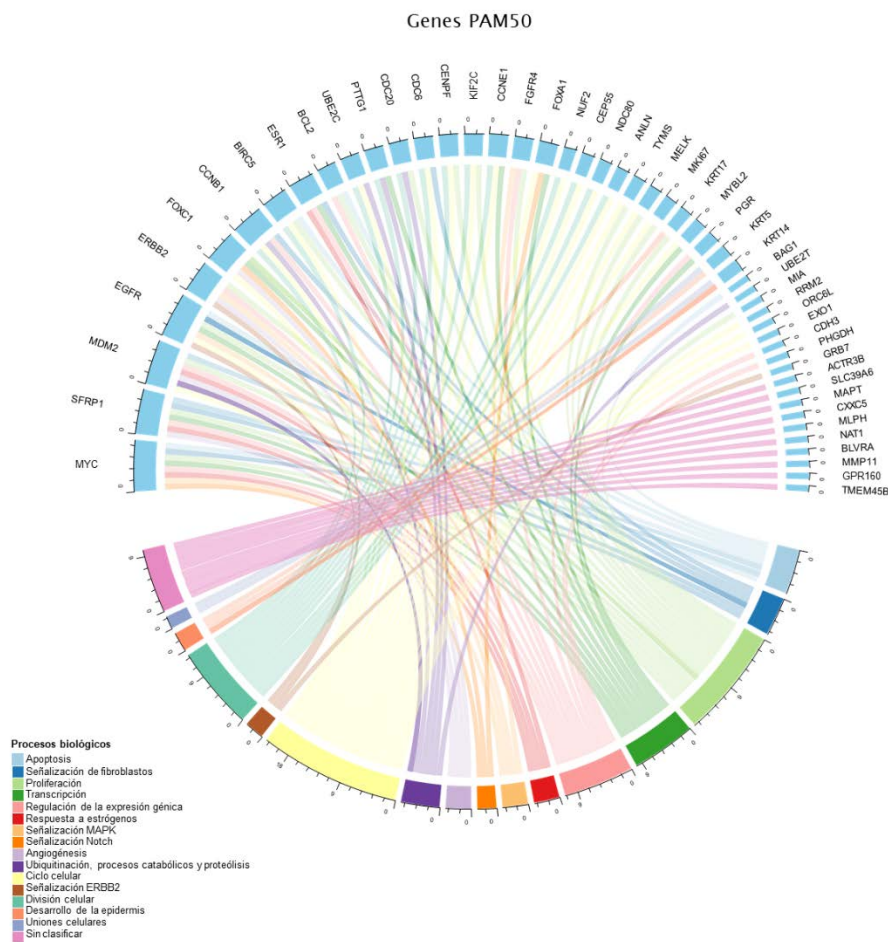


Figura 6. Genes incluidos en la firma PAM50. En la parte superior del *circus plot* se muestran los genes que se incluyen el panel PAM50. En la parte inferior se muestran los procesos biológicos a los que pertenecen dichos genes.

3.3.2.2. Panel *PanCancer pathways*: Análisis de expresión diferencial, correlación entre genes diferencialmente expresados y enriquecimiento en rutas biológicas.

Continuando con la tecnología nCounter DX Analysis System® explicada en el apartado anterior, se evaluaron los cambios encontrados entre los diferentes tiempos según los diferentes objetivos del presente trabajo. Con el panel PanCancer Pathways® se determinó la expresión de 730 genes (Tabla 1) involucrados en las 13 vías principales asociadas a cáncer, más la expresión de 40 genes como genes *housekeeping* (usados como control positivo dado a su expresión constitutiva bajo cualquier condición). Información más detallada del panel puede ser consultada en <https://www.nanostring.com>. Para poder llevar a cabo la cuantificación de la expresión génica, usamos 50ng de RNA de las muestras de las pacientes y dicho RNA debía tener un valor de la ratio 260nm/280nm por encima del 1.7, asegurándonos la calidad de este. Todas aquellas muestras que no tuvieran suficientes nanogramos de RNA o la ratio 260nm/280nm estuviera por debajo de 1.7 fueron descartadas.

Tanto la normalización de los resultados, como el análisis de expresión diferencial (AED) y el enriquecimiento en vías o rutas biológicas fue llevado a cabo con el software Nanostring nCounter nSolver™ 4.0, usando el módulo Nanostring Advance Analysis Module 2.0. El módulo avanzado de nSolver™ 2.0 es una herramienta basada en el lenguaje de programación “R” proporcionada por Nanostring technologies específica para el tipo de datos que el sistema nCounter® DX Analysis System genera, es decir, archivos .RCC. Esta herramienta es de libre acceso y realiza el control de calidad de las muestras, la normalización de los datos, el AED y el enriquecimiento en set de genes asociados a rutas biológicas.

Tabla 1. Genes incluidos en el panel PanCancer Pathways.

730 genes incluidos en el PanCancer Pathways Panel

ABL1	BIRC7	CASP8	CHEK1	DDIT4	EYAI	FIGF	GRIA3	IBSP	IL3	LAMA1	MAPT	NFKBIA	PDGFRB	PPARGC1A	RAD51	SIN3A	TBL1XR1	UTY
ACVR1B	BMP2	CASP9	CHEK2	DIK1	EZH2	FLNA	GRIN1	ID1	IL3RA	LAMA3	MCM4	NFKB1Z	PGF	PPF2CB	RAD52	SIRT4	TCF3	VEGFA
ACVR1C	BMP4	CBL	CHUK	DIK2	FANCA	FLNC	GRIN2A	ID2	IL5RA	LAMA5	MCM5	NGF	PHF6	PPP2R1A	RAF1	SIX1	TCF7L1	VEGFC
ACVR2A	BMP5	CBLC	CIC	DIK4	FANCB	FLT1	GRIN2B	ID4	IL6	LAMB3	MCM6	NGFR	PIK3CA	PPP2R2B	RASA4	SKP1	TCL1B	VHL
AKT1	BMP6	CCNA1	CLCF1	DLL1	FANCC	FLT3	GSK3B	IDH1	IL6R	LAMB4	MCM7	NKD1	PIK3CB	PPP2R2C	RASAL1	SKP2	TET2	WEE1
AKT2	BMP7	CCNA2	CNTFR	DLL3	FANCE	FN1	GIT2H3	IDH2	IL7	LAMB2	MDC1	NODAL	PIK3CD	PPP3CA	RASCRF1	SMAD2	TDFP1	WHSC1L1
AKT3	BMP8A	CCNB1	COL11A1	DLL4	FANCF	FOS	CZMB	IFNA17	IL7R	LAMB3	MDM2	NODC	PIK3CC	PPP3CB	RASGRF2	SMAD3	TGFB1	WHSC1L1
ALK	BMPRI1B	CCNB3	COL11A2	DNMT1	FANCC	FOSL1	H2AFX	IFNA2	IL8	LAT	MECOM	NO63	PIK3R1	PPP3CC	RASCRP1	SMAD4	TGFB2	WIF1
ALKB2	BRAF	CCND1	COL1A1	DNMT3A	FANCL	FOXO1	H3F3A	IFNA4	INHBA	LEFT1	MED12	NOTCH1	PIK3R2	PPP3R1	RASCRP2	SMAD9	TGFB3	WNT10A
ALKB3	BRAF	CCND2	COL1A2	DTX1	FAS	FOXO4	H3F3C	IFNG	INHBB	LEFTY1	MEN1	NOTCH2	PIK3R3	PPP3R2	RB1	SMARCA4	TGFB3	WNT10B
AMER1	BRCA1	CCND3	COL24A1	DTX3	FASLG	FST	HDAC1	IGF1	IRAK2	LEFTY2	MET	NOTCH3	PIK3R5	PRDM1	RBX1	SMARCB1	THBS1	WNT11
AMH	BRCA2	CCNE1	COL27A1	DTX4	FBXW7	FUBP1	HDAC10	IGF1R	IRAK3	LEP	NPM1	NPY2	PIK3R4	PRKAA2	RELA	SMC1A	THBS4	WNT16
ANGPT1	BRIPI1	CCNE2	COL2A1	DIUSP10	FEN1	FUT8	HDAC11	IGFBP3	IRS1	LEPR	MGMT	NPY2	PITX2	PRKACA	REIN	SMC1B	THBS4	WNT2
APC	CT9orf40	CCNO	COL3A1	DIUSP2	FGF1	FZD10	HDAC2	IKBKB	ITGA2	LFNG	MLF1	NPY2	PITX2	PRKACB	RET	SMC3	TIAM1	WNT2B
APH1B	CACNA1C	CCR7	COL4A3	DIUSP4	FGF10	FZD2	HDAC4	IKBKG	ITGA3	LIF	MLH1	NR4A3	PLA1A	PRKACC	RFC3	SMO	TLR2	WNT3
AR	CACNA1D	CD14	COL4A4	DIUSP5	FGF11	FZD3	HDAC5	IL10	ITGA6	LIFR	MLL3	NRAS	PLA2G10	PRKAR1B	RFC4	SOCS1	TLR4	WNT4
ARID1A	CACNA1E	CD19	COL4A5	DIUSP6	FGF12	FZD7	HDAC6	IL11	ITGA7	LIG4	MLL7	NRAS	PLA2G2A	PRKAR2A	RHOA	SOCS2	TLX1	WNT5A
ARID1B	CACNA1G	CD40	COL4A6	DIUSP8	FGF13	FZD8	HELLS	IL11RA	ITGA8	LRP2	MMP3	NTF3	PLA2G3	PRKAR2B	RINI	SOCS3	TNFRSS2	WNT5B
ARID2	CACNA1H	CD414A	COL5A1	E2F1	FGF14	FZD9	HES1	IL12A	ITGA9	LTBP1	MMP7	NTHL1	PLA2G4A	PRKCA	RNF43	SO51	TNC	WNT6
ARNT2	CACNA2D1	CD414B	COL5A2	E2F5	FGF16	GADD45A	HES5	IL12B	ITGB3	MAD2L2	MMP9	NTRK1	PLA2G4C	PRKCB	RPA3	SO52	TNF	WNT7A
ASXL1	CACNA2D2	CD225A	COL6A6	EFNA1	FGF17	GADD45B	HGF	IL12RB2	ITGB4	MAML2	MINA1	NTRK2	PLA2G4E	PRKCG	RP527A	SOST	TNFAIP3	WNT7B
ATM	CACNA2D3	CD225B	COMP	EFNA2	FGF18	GADD45C	HHEX	IL13	ITGB6	MAP2K1	MPL	NUMBL	PLA2G4F	PRKDC	RP56KA5	SOX17	TNFRSF10A	WTT1
ATR	CACNA2D4	CD225C	CREB3L1	EFNA3	FGF19	GAS1	HHIP	IL13RA2	ITGB7	MAP2K2	MPO	NUPR1	PLA2G5	PRKX	RP56KA6	SOX9	TNFRSF10B	XPA
ATRX	CACNB2	CD26	CREB3L3	EFNA5	FGF2	GATA1	HIST1H3B	IL15	ITGB8	MAP2K4	MSH2	OSM	PLAT	PRL	RRA52	SP1	TNFRSF10C	XRCC4
AXIN1	CACNB3	CD27	CREB3L4	EGF	FGF20	GATA2	HIST1H3C	IL19	JAG1	MAP2K6	MSH6	PAK3	PLAU	PRLR	RUNX1	SPOP	TNFRSF10D	ZAK
AXIN2	CACNB4	CDH1	CREB5	EGFR	FGF21	GATA3	HIST1H3H	IL17A	JAG2	MAP3K1	MTOR	PAK7	PLCB1	PRMT8	RUNX1T1	SFP1	TNFRSF10E	ZEB1
B2M	CACNG1	CDK2	CREBBP	EIF4EBP1	FGF22	GDF6	HMCA1	IL1B	JAK1	MAP3K12	MUTYH	PAX3	PLCB4	PROM1	RXRG	SPRY1	TNN	ZEB3
BAD	CACNG4	CDK4	CRLF2	ENDOG	FGF23	GHR	HMCA2	IL1R1	JAK2	MAP3K13	MYB	PAX5	PLCE1	PTCH1	SETBP1	SPRY2	TNR	ZIC2
BALF3	CACNG6	CDK6	CSF1R	EP300	FGF3	GLI1	HNF1A	IL1R2	JAK3	MAP3K14	MYC	PAX8	PLCG2	PTCRA	SETD2	SPRY4	TP53	
BAMBI	CALML3	CDKN1A	CSF2	EPHA2	FGF4	GLI3	HOXA10	IL1RAP	JUN	MAP3K5	MYCN	PBRM1	PLD1	PTEN	SF3B1	SRSF2	TPO	
BAP1	CALML5	CDKN1B	CSF3	EPO	FGF5	GNA11	HOXA11	IL20RA	KAT2B	MAP3K8	MYD88	PBX1	PMIL	PTPN11	SFN	SSX1	TRAF7	
BAX	CALML6	CDKN1C	CSF3R	EPOR	FGF6	GNAQ	HOXA9	IL20RB	KOM5C	MAPK1	NASP	PBX3	POLB	PTPN5	SFRP1	STAG2	TSC1	
BCL2	CAMK2B	CDKN2A	CTNNB1	ERBB2	FGF7	GNA5	HPGD	IL22RA1	KDM6A	MAPK10	NBN	PCK1	POLD1	PTPRR	SFRP2	STAT1	TSHR	
BCL2L1	CAPN2	CDKN2B	CUL1	ERCC2	FGF8	GNG12	HRA5	IL22RA2	KIT	MAPK12	NCOR1	PCNA	POLD4	PITTC2	SFRP4	STAT3	TSLP	
BCOR	CASP10	CDKN2C	CXIC4	ERCC6	FGF9	GNG4	HSP90B1	IL23A	KITLG	MAPK3	NF1	PCGFA	POLE2	RAC1	SGK2	STAT4	TSPAN7	
BDNF	CASP12	CEBPA	DAXX	ETV1	FGFR2	GNG7	HSPA1A	IL23R	KLFA	MAPK8	NF2	PCGFB	POLR2D	RAC2	SHC1	STK11	TTK	
BID	CASP3	CEBPE	DOB2	ETV4	FGFR3	GPC4	HSPA6	IL2RA	KMT2C	MAPK8IP1	NFATC1	PCGFC	POLR2H	RAC3	SHC2	STMN1	U2AF1	
BIRC3	CASP7	CHAD	DDIT3	ETV7	FGFR4	GRB2	HSPB1	IL2RB	KRA5	MAPK9	NFKB1	PCGFR	POLR2J	RAD21	SHC3	SUV39H2	UBE2T	



3.3.2.2.1. Control de calidad y normalización de los datos

Los datos en bruto de las muestras procesadas tenían que pasar por unos controles de calidad para comprobar que el proceso había sido correcto. Los parámetros incluidos fueron: *Imaging QC* (nitidez de la foto tomada por la cámara), *Binding Density QC* (densidad de los códigos de colores captados por la cámara), *Positive Control Linearity QC* (linealidad entre las diferentes concentraciones de los controles positivos) y *Positive Control Limit of Detection QC* (detección de una concentración mínima de los controles positivos). Todos estos parámetros debían estar entre unos valores concretos (Figura 7) definidos por el fabricante. Todas las muestras fueron normalizadas usando la media geométrica de la expresión de los genes *housekeeping*, y aunque esta normalización podía ser realizada de manera automática, se decidió hacer una normalización manual de los datos para así mantener en todos los análisis de expresión génica para todos los capítulos del presente trabajo los mismos genes *housekeeping* (Tabla 2).

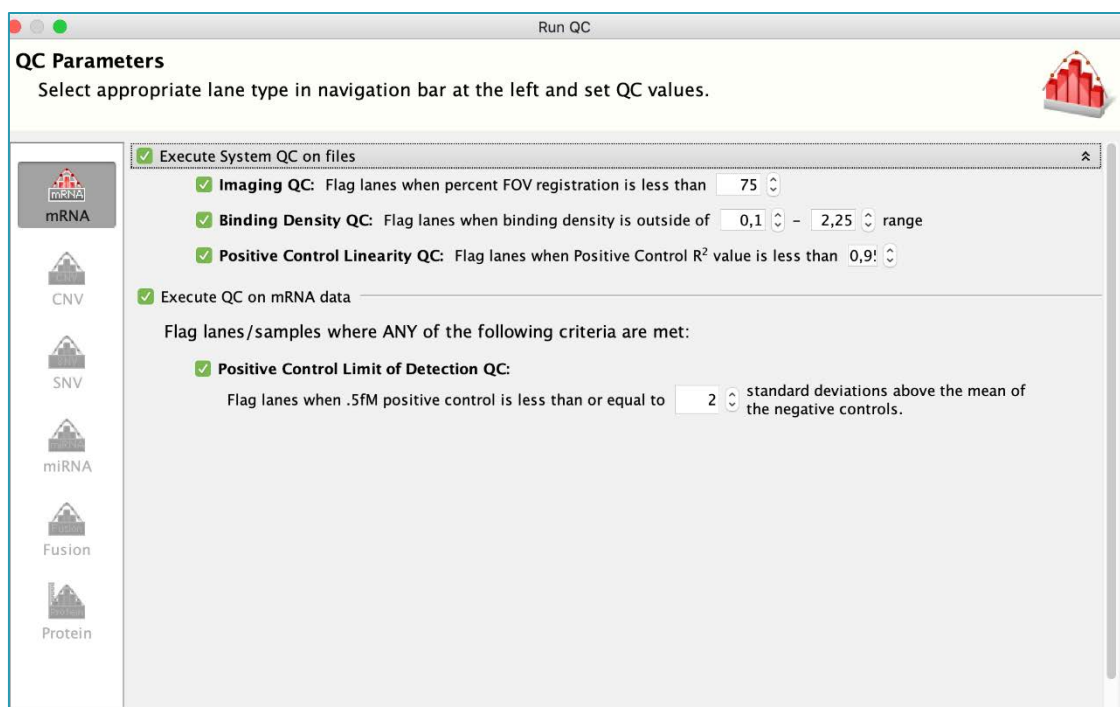


Figura 7. Imagen de la herramienta nSolver™ con los parámetros en los controles de calidad y los valores límites para que las muestras sean incluidas en los posteriores análisis.

Tabla 2. Genes *housekeeping*. Listado de los 40 genes incluidos en el panel *PanCancer Pathways* que se toman para la normalización de los datos.

Genes <i>housekeeping</i>			
ACAD9	DNAJC14	NOL7	TMUB2
AGK	EDC3	NUBP1	TRIM39
AMMECR1L	EIF2B4	PIAS1	TTC31
C10orf76	ERCC3	PIK3R4	USP39
CC2D1B	FCF1	PRPF38A	VPS33B
CNOT10	FTSJ2	RBM45	ZC3H14
CNOT4	GPATCH3	SAP130	ZKSCAN5
COG7	HDAC3	SF3A3	ZNF143
DDX50	MRPS5	SLC4A1AP	ZNF346
DHX16	MTMR14	TLK2	ZNF384

3.3.2.2.2. Análisis de expresión diferencial en T1 entre pCR y no-pCR (Capítulo I)

Como nuestro objetivo en cuanto a los análisis de expresión génica era conocer aquellos genes (Tabla 1) que se comportan de forma diferente entre dos situaciones concretas, la aproximación más interesante era realizar un AED que tendría como resultado los genes diferencialmente expresados (DE) en el caso de que los haya. El AED aplica, preferencialmente, el método estadístico óptimo por cada gen según la distribución de la variable, pudiendo ser 1) Modelo negativo mixto binomial, 2) Modelo negativo simplificado binomial o 3) Modelo Logarítmico Lineal, en ese orden. El p-valor devuelto por la prueba estadística fue ajustado por la prueba FDR (*false discovery ratio*) por el método Benjamini-Hochberg. Para calcular el logaritmo en base 2 del *Fold Change* (Log₂ FC) o la cantidad de cambio y la dirección de este (sobreexpresión o expresión a la baja) entre los genes de la situación de pCR y no-pCR, se determina la log-ratio entre las medidas de ambas situaciones. Estos valores pueden ser mayores de 0 (sobreexpresado en pCR con respecto a no-pCR) o menores de 0 (expresados a la baja en pCR con respecto a no-pCR).

3.3.2.2.3. Análisis de expresión diferencial en el análisis longitudinal (capítulo II)

El análisis de expresión diferencial entre diferentes tiempos requiere de unos parámetros específicos no incluidos en el software nSolver™. Para poder ver aquellos cambios entre los diferentes escenarios en los que nos encontramos (Pre_T1vsRD_T2, Pre_T1vsMtx_T3 y RD_T2vsMtx_T3) usamos el programa Deseq2 de R (4.0.2) que

permite hacer pruebas entre muestras pareadas mediante el diseño de una fórmula que modela las diferencias a través del tiempo con un *Likelihood Ratio Test* (LRT). Como partimos de los datos procedentes de la tecnología nCounter® de Nanostring, se adaptaron para que fueran analizables en el Deseq2 con el método *removing unwanted variation* (RUV) (86).

3.3.2.2.4. Enriquecimiento en rutas biológicas

Aplicando un análisis de componentes principales, el módulo avanzado 2.0 de la herramienta de análisis nSolver™ es capaz de generar un *pathway-score*, con un análisis de componentes principales, que trata de resumir toda la aportación que hacen los genes que estamos sometiendo a estudio a una ruta biológica concreta. En el *pathway-score* se resumen los datos de los genes de una ruta en una única puntuación. Las puntuaciones de las rutas se calculan como el primer componente principal de la expresión normalizada de los genes de las rutas. A la media de este *pathway-score* se le aplica el test *t de student* para ver si existen diferencias significativas entre dos situaciones concretas que nos interesen. La dirección positiva o negativa de la diferencia entre ambas situaciones está relacionada con la expresión de los genes que componen esa ruta. En el panel *PanCancer Pathways* se contemplan las 13 rutas principales más importantes relacionadas con cáncer (Tabla 3, Figura 8).

Tabla 3. Rutas biológicas que engloban a los 730 genes incluidos en el panel

Rutas biológicas incluidas en el <i>PanCancer Pathways</i>	
Notch	MAPK
APC (Wnt)	JAK/STAT
HedgeHog	PI3K
Chromatin Modification	RAS
Transcriptional Regulation	Cell Cycle
DNA Damage Control	Apoptosis
TGF-β	

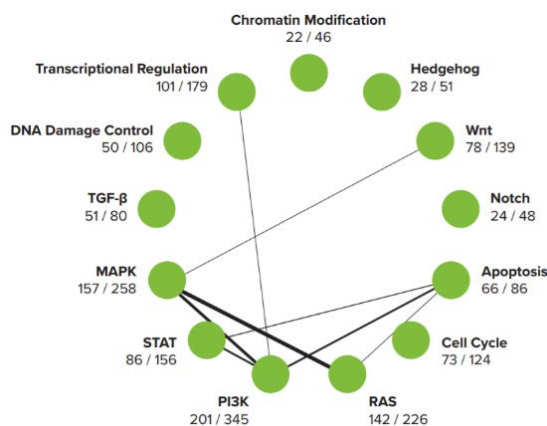


Figura 8. Vías incluidas en el panel *PanCancer Pathways* de nCounter®. Los círculos representan de cada una de las 13 vías canónicas el número de genes seleccionados para el panel en relación con el número total de genes conocidos para cada vía identificado por KEGG. Las líneas han sido dibujadas para mostrar la relación de los genes que pertenecen a diferentes vías, con el grosor de la línea se determina el número de genes compartidos.

3.3.3. Predicción de pCR en T1

Tal y como hemos especificado en la introducción, actualmente hay variables clinicopatológicas ya conocidas que predicen pCR, sin embargo, ninguna de ellas se posiciona como un claro marcador de la respuesta a la NAC en la comunidad científica. Con el objetivo de mejorar la capacidad predictiva de pCR de los biomarcadores ya conocidos, en las variables usadas para el diseño de modelos predictivos hemos incluido tanto las variables clinicopatológicas predictoras de pCR ya conocidas, como otras variables también clinicopatológicas que nos parecían interesantes y los genes diferencialmente expresados del panel *PanCancer Pathways* entre aquellos pacientes con pCR y no-pCR. Para asegurarnos de que la información de los genes DE aportaban valor predictivo a las variables ya conocidas, o incluso podían aportar más información por ellos mismos, se han utilizado 4 set de datos; por un lado, el **set CLIN** que incluye las variables clinicopatológicas incluye los biomarcadores ya conocidos, el **set GENEX** que incluye sólo la información de los genes diferencialmente expresados, el **set CLIN_GENEX** que incluye tanto la información clínica como la información genética y el **set PAM50**, que incluye únicamente el dato de los subtipos determinados por PAM50.

Con el fin de identificar modelos predictivos que pudieran ser específicos de las características más importantes definitorias de subtipos en cáncer de mama (que a su vez se encuentran relacionados con la obtención de pCR o en la supervivencia del paciente), como son la sobreexpresión de HER2 o el estatus de ER, hemos subdividido las pacientes en 3 grupos. **Cohorte_*Allsubtypes***, todas las pacientes con independencia de si tiene sobreexpresión de HER2 o el estatus de ER. **Cohorte_*HER2-***, únicamente aquellas pacientes que no sobreexpresan el gen HER2+ (con independencia del estatus de ER) y la **Cohorte_*ER+*** para aquellas pacientes con el estatus de ER positivo y sin sobreexpresión del gen HER2. Por otra parte, ha sido imposible obtener cohorte suficientemente grande positiva para HER2 y otra ER-/HER2- debido al bajo número de pacientes con estas características.

Todos los análisis de los datos obtenidos y evaluados como predictores de pCR han sido desarrollados en el lenguaje de programación R (versión 3.6).

3.3.3.1. Análisis Univariante

Para determinar la relación de cada una de las variables a estudio (variables independientes) por separado con la variable problema (variable dependiente), hemos realizado un análisis univariante basado en una regresión logística (RL). Como medida de bondad de ajuste del modelo usamos el LRT. EL LTR nos permitirá comparar un modelo simple con un modelo más complejo por comparación de las diferencias en la desviación de ambos modelos. En el caso del análisis univariante, el modelo simple es el modelo sin predictores y el modelo complejo es con una sola variable predictiva. Consideraremos las variables predictivas como estadísticamente significativas cuando el p-valor del LRT sea menor de 0.05. Las variables a estudio (variables independientes) serán las variables clínico-patológicas recogidas y los genes diferencialmente expresados entre las pacientes que obtienen pCR y las que no, mientras que la variable problema o dependiente será conseguir pCR.

3.3.3.2. Análisis multivariante

El análisis univariante nos da una visión de cómo las características o variables independientes, cada una por separado, afectan a la variable dependiente. Para

determinar si un conjunto de variables independientes (características clinicopatológicas más los genes diferencialmente expresados entre pCR y no-pCR) pueden determinar o predecir de manera combinada a la variable dependiente (pCR), realizamos un análisis multivariante. Para ello, se ha empleado un modelo Lasso con doble validación cruzada. En la **validación interna se realiza la selección de variables con Lasso** y el ajuste de los hiperparámetros mediante una validación con 5 cajas. Para la **validación externa de la selección de variables se usó *Leave-one out cross-validation* (LOOCV)** ya que es la validación que mejor se ajusta a cohortes heterogéneas y con pocos pacientes, como es nuestro caso. Con la métrica **AUC** se evaluó la selección de características y para hacer más robustos nuestros resultados este proceso se repitió 1000 veces. La importancia de cada característica dentro del modelo fue medida como el número de veces que fue seleccionada durante cada interacción.

El **desbalanceo de datos de las categorías de una variable** puede afectar al rendimiento del aprendizaje de los algoritmos. La presencia de una clase infrarrepresentada (en nuestra cohorte hemos visto dicho desbalanceo entre los grupos pCR y no-pCR) conduce a la formación de modelos inútiles, por lo que se ha usado el **algoritmo SMOTE** que maneja la distribución de clases desequilibradas sobremuestreando la clase minoritaria. Esta técnica ha sido aplicada cuando el porcentaje de la clase minoritaria es inferior al 30%.

Una vez establecidos los modelos predictivos de pCR, con los diferentes sets de datos y en los diferentes grupos de pacientes, para sus **evaluaciones ha sido usada, de nuevo, la métrica AUC con validación cruzada**, que mide la capacidad de predicción. AUC hace referencia al área comprendida debajo de la curva ROC, y la curva ROC es una medida de la especificidad (proporción de pacientes negativos para la pCR correctamente identificados) y la sensibilidad (proporción de pacientes positivos para la pCR correctamente identificados) de nuestro modelo. El valor predictivo positivo (VPP) de nuestro modelo mide la proporción de pacientes que han sido positivos para pCR y efectivamente lo son y el valor predictivo negativo (VPN) mide la proporción de pacientes que han sido pronosticados como no-pCR y lo son.

Para la **validación de nuestros modelos**, hemos utilizado cohortes externas. Todos los modelos obtenidos con los diferentes sets de datos (CLIN, CLIN_GENEX y GENEX) y con los diferentes grupos de pacientes según HER2 y ER (Cohorte_*Allsubtypes*, Cohorte_HER2+ y Cohorte_ER+) han sido ejecutados con los datos de cohortes externas de estudios anteriores donde las pacientes presentaban el mismo esquema de tratamiento y tenían tanto la información clínico-patológica como de expresión génica necesaria publicada en repositorios públicos. Dos estudios han sido encontrados que cumplan con los datos que nuestros modelos de predicción requieren, un estudio de Tabchy y colaboradores (2010) (87) (acceso a los datos en GEO DataSet: GSE20271) y otro de Hatzis y colaboradores (2011) (88) (acceso a los datos en GEO DataSet: GSE25066). Como la distribución de pacientes con diferentes estatus de HER2 y ER no es homogénea en ninguno de los dos estudios, los modelos predictivos definidos con los diferentes sets de datos se han validado en los grupos de pacientes según las características de HER2 y ER sin tener en cuenta el estudio del que procede. De esta manera, vamos a diferenciar entre nuestra cohorte de estudio donde los modelos predictivos han sido generados (**cohortes_*Discovery***), y las cohortes externas a nuestro estudio donde hemos validado los modelos predictivos generados que denominamos cohorte_*Discovery* (**cohortes_*Validation***) Figura tal. Para la evaluación de los modelos en las cohortes_*Validation* se ha usado la métrica AUC con validación cruzada, de la misma forma que en las cohortes_*Discovery* (Figura 9).

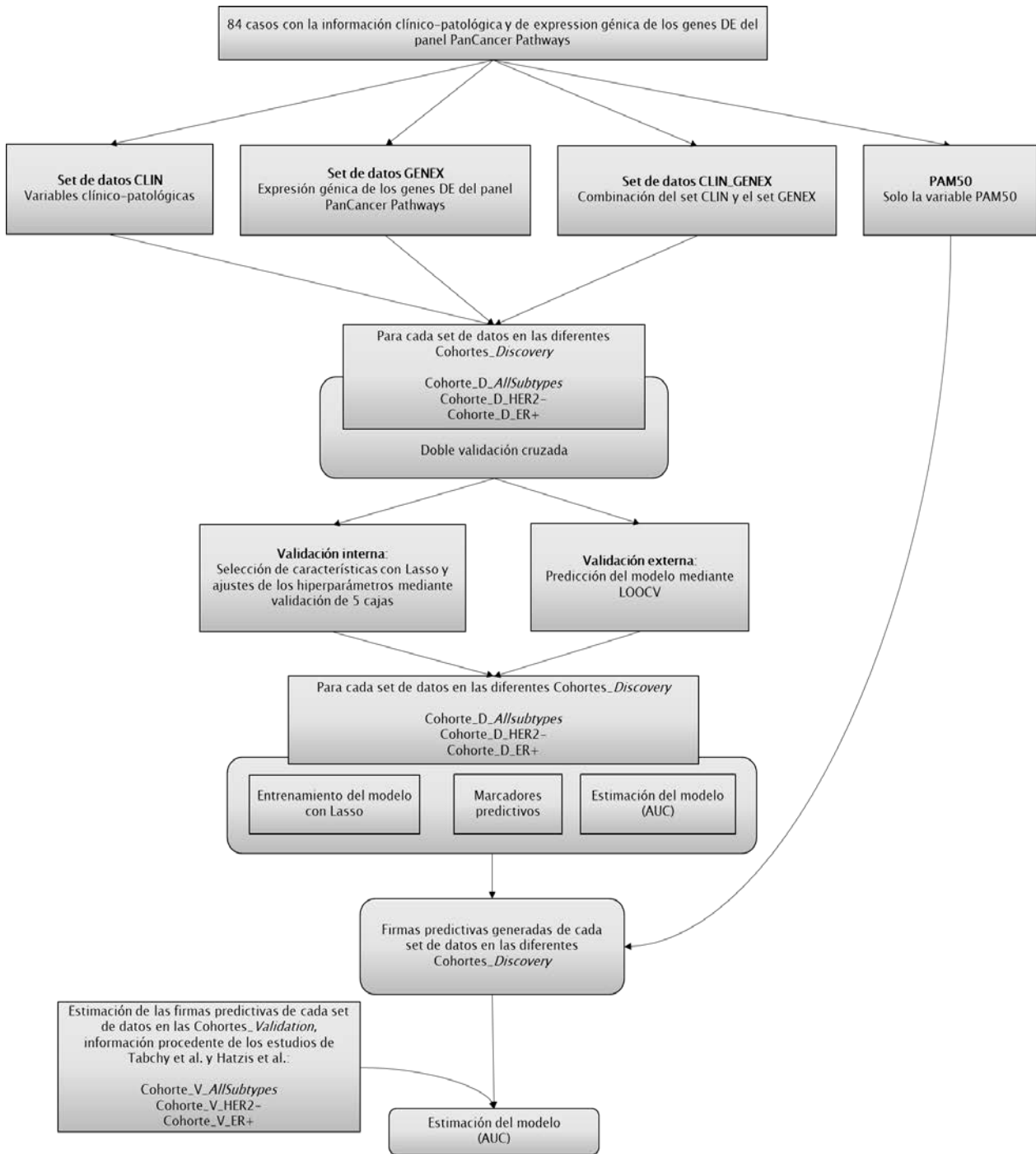


Figura 9. Flujo de trabajo para la determinación de los modelos predictivos. Con los sets de datos CLIN, GENEX y CLIN_GENEX se han determinado las variables más importantes para la determinación de pCR en las pacientes de las Cohortes_Discovery con una doble validación cruzada, una validación interna con Lasso y una validación externa por LOOCV. Paralelamente se ha incluido un modelo solo con la variable PAM50. De nuevo, Lasso ha sido el método usado para el entrenamiento de los modelos y la métrica para su estimación, AUC. Una vez las firmas han sido generadas, han sido probadas en las diferentes cohortes tanto en las Cohortes_Discovery como en las cohortes_Validation usando la métrica de AUC para su estimación. El modelo PAM50 solo ha sido probado en las Cohortes_Discovery.

3.4. ESTUDIO DE LAS ALTERACIONES GENÉTICAS

3.4.1. Aislamiento del DNA desde el tejido embebido en parafina.

Para la extracción del DNA de las muestras FFPE se llevó a cabo una macro-disección del tejido tumoral previamente delimitada por el patólogo. La purificación del DNA se llevó a cabo con el kit *GeneRead DNA FFPE* (QIAGEN). Para determinar la pureza y la concentración del DNA aislado se usaron dos métodos: espectrofotometría de bajo volumen (Nanodrop™, Thermo Scientific), donde la ratio entre 260nm/280nm (A260/280) determina la pureza de los ácidos nucleicos con respecto a las proteínas (se han dado por válidos valores próximos a 1.8) y espectrofotometría con fluorocromos específicos de DNA de cadena doble (QUBIT™, Invitrogen). Para identificar una posible contaminación por sales o alcoholes derivados del proceso de aislamiento también se tiene en cuenta la ratio 260nm/230nm (A260/230) proporcionada por Nanodrop™, las muestras debían estar en torno a 2 para ser incluidas.

3.4.2. Genes y secuenciación.

Se han secuenciado un total de 52 genes implicados en la biología del cáncer de mama, quedan especificados en la Tabla 4.

La elaboración de las librerías para la secuenciación de los genes diana se ha llevado a cabo con la química de AmpliSeq™ cuyo procesamiento de la muestra solo necesita de 1 ng de DNA, y cuenta con la flexibilidad suficiente como admitir muestras fijadas y embebidas en parafina, como es el caso de este estudio. La secuenciación de las muestras se ha realizado con la plataforma de illumina® MiSeq™ basada en la tecnología de secuenciación por síntesis (Figura 10).

Tabla 4. Genes incluidos en el panel de secuenciación.

ABCB1	CYP1B1	KRAS	PIK3R1
AKT1	EGFR	Malat1	PLEkhs1
ARID1A	ERBB2	MAP2K4	POLE
ATM	ERCC2	MAP3K1	PTEN
AURKA	ERCC5	MCL1	RB1
BRCA1	ESR1	Med23	RECQL4
BRCA2	FANCA	Mllt4	tbc1d12
CCND1	FANCD2	MSH6	TBX3
CCND3	FGFR1	MYC	TP53
CCNE1	FOXA1	NCOR1	
CDH1	GATA3	Neat1	
CDK6	JAK2	NF1	
CDKN2A	KMT2C	PALB2	
CDKN2B	KMT2D	PI3KCA	

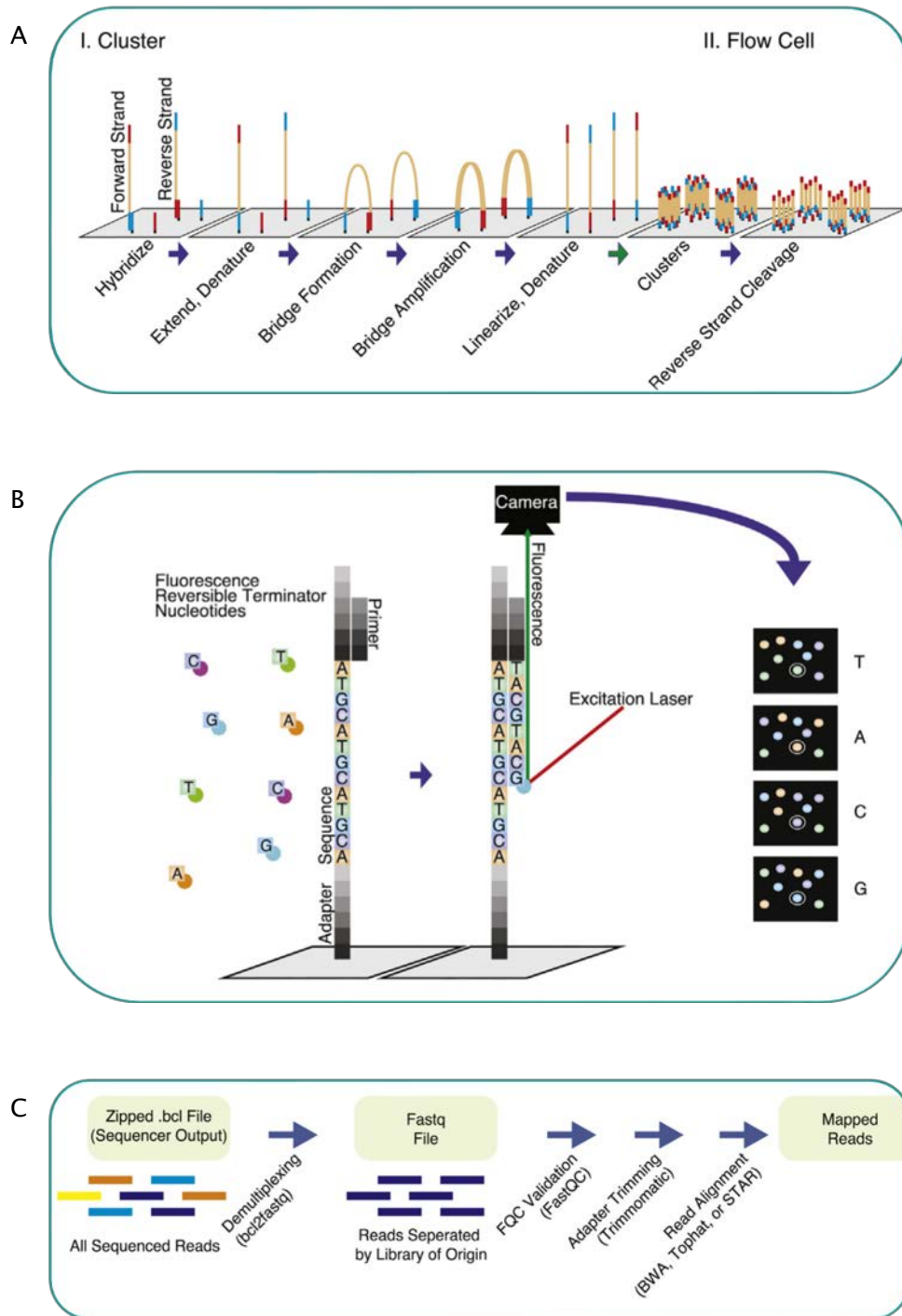


Figura 10. Flujo de trabajo para la secuenciación de los genes diana. A. Los fragmentos desnaturalizados de las librerías fluyen a través de la *flowcell* uniéndose a los adaptadores que hay presente en la superficie, que son complementarios a los adaptadores de nuestras muestras. Los fragmentos complementarios son extendidos y amplificados con una PCR-puente, tras esto serán desnaturalizados formando clústeres desde un mismo fragmento. B. Dichos fragmentos son secuenciados utilizando nucleótidos terminadores reversibles. Los pares de bases se identifican después de la excitación del láser y la detección de la fluorescencia. C. Los datos brutos de demultiplexan en librerías individuales y se evalúa la calidad. Se eliminan las lecturas de los adaptadores para eliminar el ruido técnico.

3.4.3. Análisis bioinformático de las variantes encontradas a nivel de DNA: Mutaciones.

El análisis bioinformático se realizó a través de la plataforma *Genome one* (Dreamgenics, S. L.), certificada con la marca IVD/CE de productos sanitarios para diagnóstico in vitro de acuerdo con la legislación vigente (Número de Licencia: 7157-PS). El procesamiento integra los siguientes pasos: Los archivos crudos FASTQ fueron evaluados usando los controles de calidad de FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y *Trimmomatic* (89) para la eliminación de bases, adaptadores y otras secuencias de baja calidad. Cada FASTQ se alineó frente a la versión GRCh37/hg19 del genoma humano de referencia mediante BWA-mem (90) y la generación de los archivos BAM ordenados se hizo con SAMtools (91). Al tratarse de un panel de amplicones, no se eliminaron lecturas duplicadas. SNVs e *indels* fueron identificadas mediante la combinación de VarScan 2 (92) y un algoritmo de llamada a variantes propio de Dreamgenics S.L. Las variantes se anotaron utilizando distintas bases de datos con información funcional (RefSeq, Pfam), poblacional (dbSNP, 1000 *Genomes*, ESP6500, ExAC, gnomAD), de predicción de impacto funcional *in silico* (dbNSFP, dbSNV), e información relacionada con el cáncer (COSMIC, ICGC).

En cuanto a la **priorización de variantes**, para la selección de variantes potencialmente somáticas en cada muestra se aplicaron los siguientes filtros generales:

- Variantes con presencia en COSMIC
- Predicción de efecto en proteína
- Frecuencia alélica poblacional inferior a 0.05
- Cobertura total en la posición afectada superior a 50X
- Frecuencia alélica de la variante en la muestra superior a 0.1

En el **análisis comparativo**, las diferencias entre distintos tiempos se analizaron con el software GraphPad Prism® y se realizó una prueba de *Fisher* para comprobar la significancia estadística.

El **meta-análisis** con la integración de variables clínicas para predecir la variable de interés, en este caso la recurrencia a distancia, entre T1 y T2 se hizo mediante la aproximación de *Machine Learning* conocida como *Random forest*.

3.5. ANÁLISIS DE SUPERVIVENCIA

Con la intención de comprobar si alguno de los resultados obtenidos en las comparativas realizadas (ya sea para el objetivo 1, 2 o 3), con la información clínico-patológica, la expresión génica y el análisis del DNA, pudiera estar relacionado con el fallecimiento de la paciente, hemos llevado a cabo un análisis de supervivencia. La fecha de la muerte con enfermedad (MCE) del paciente será la fecha del evento que queremos analizar. Usando el lenguaje de programación R (versión 3.6) con el paquete *Survival*, estimamos la función de supervivencia, así como la curva de Kaplan-Meier para calcular la supervivencia global y los intervalos de confianza. Para los datos categóricos se ha usado el test log-rank para comparar las curvas de Kaplan-Meier de diferentes grupos. En el caso de la expresión génica, se ha categorizado en base a los cuartiles en baja y alta expresión. Además, también hemos llevado a cabo un modelo de regresión de Cox. La estimación de los parámetros se ha realizado por *partial likelihood*. El análisis univariante y multivariante ha sido realizado para las variables clínico-patológicas, la información de la expresión génica y los datos del análisis de DNA. Para cada modelo, se ha determinado la significancia de la variable con el *Likelihood Ratio test*. En el caso del modelo multivariante, la selección de variables se ha realizado por el algoritmo *stepwise*. Se han usado dos criterios de métrica, BIC (*Bayesian Information Criteria*) y AIC (*Akaike Information Criteria*). Finalmente, se ha verificado las restricciones impuestas en los modelos de Cox: asunción de riesgos proporcionales, inspección de los *deviance residuals* como método de detección de *outliers*, empleo de *splines* penalizados en los residuos de *Martingale* para identificar la relación entre cada variable y la razón logarítmica de riesgo y, finalmente, la identificación de puntos de mayor influencia a través de los residuos de *Schoenfel*.

3.6. LIMITACIONES DE LOS ESTUDIOS

De forma general, la principal limitación que alberga el proyecto de donde se han desarrollado los dos estudios o capítulos que forman la presente tesis es el contemplar solo un número concreto de genes tanto en el análisis de expresión génica como en los estudios de secuenciación. Al centrarnos en aquellos genes con una relación previa con cáncer, los resultados tienen una relación directa con la enfermedad, pero por otro lado dejamos fuera la posible determinación de nuevos genes que pueden no haber sido identificados aún.

En cuanto al capítulo I, la escasez de material de las biopsias diagnósticas nos ha llevado a centrarnos en el estudio de la expresión génica y no poder desarrollar un estudio amplio en la determinación de mutaciones. Otra limitación que encontramos en este capítulo es a la hora de determinar predictores de pCR en los diferentes subtipos de cáncer de mama, carecemos de cohortes de pacientes únicamente con subtipo HER2+ y pacientes TN en las que tenga un número suficiente de pacientes para determinar un posible modelo predictivo.

Por otra parte, en el capítulo II, cuando se estudia las similitudes/diferencias entre el tumor primario o la enfermedad residual con respecto a la primera metástasis, no hemos diferenciado entre los órganos donde se desarrollado la metástasis, sino que hemos determinados marcadores generales de metástasis, es posible que la biología de cada órgano influya considerablemente en la biología propia del tumor metastásico. Tampoco se ha considerado el tratamiento adyuvante que la paciente sigue o no tras la cirugía, ni los tiempos que definen los pasos de la historia natural del tumor.



UNIVERSIDAD
DE MÁLAGA

4. CAPÍTULO I:

Predicción de la respuesta al tratamiento Neoadyuvante

No temas a las dificultades: lo mejor surge de ellas.

Rita Levi-Montalcini

4.1. INTRODUCCIÓN Y OBJETIVOS

El tratamiento con NAC (basado en la administración de Antraciclinas+/-Taxanos, y Trastuzumab en aquellos casos con sobreexpresión de HER2) antes de la cirugía tiene un uso bastante extendido en pacientes con cáncer de mama tanto en estadios tempranos como en el escenario de la enfermedad localmente avanzada. La respuesta a la NAC es clave para tener una aproximación de la evolución del paciente después del tratamiento ya que ha sido demostrado que conseguir pCR tras la cirugía está relacionado con una mayor supervivencia global y libre de enfermedad. La respuesta al tratamiento con NAC es distinta en los diferentes subtipos de cáncer de mama, siendo los pacientes con los subtipos con las vías de proliferación más activadas, como el caso del Triple Negativo o HER2+, las más respondedoras (12) (3). Teniendo en cuenta que las pacientes que obtienen pCR en dichos subtipos están en torno al 30% (13), es vital encontrar marcadores que identifiquen aquellas pacientes que van a obtener pCR de las que no antes del tratamiento. En este escenario se han identificado tanto variables clinicopatológicas (el tamaño tumoral mamográfico, grado tumoral, subtipos determinados por inmunohistoquímica (IHQ), afectación ganglionar de la axila o % de KI67) como marcadores moleculares (subtipos moleculares definidos por PAM50, IntClust, o el modelo TRAR que se relacionan con la obtención de pCR (1)). En cuanto a las variables clínico-patológicas, el tamaño tumoral se ha descrito como una de las variables asociadas a la pCR en una cohorte distribuida en 10.3% HER2-enriched, 18,2% Luminal B, 30.6% Luminal A y 32.7% Basal-Like ($p < 0.001$). Por otra parte, aquellos tumores con un tamaño de cT1-T2 consiguen el doble de respuesta a la NAC que los tumores más grandes (cT3-T4) (2). Otra de ellas es el grado histológico tumoral; en una cohorte de 560 tumores primarios, Lips y colaboradores (2013) demostraron que un alto grado histológico tumoral (G3) predecía pCR en tumores de mama ER+/HER2- y en aquellos tumores TN tratados con NAC más sales de platino (6). Estos mismos resultados obtuvo Diaz-Redondo y colaboradores (2019) en una cohorte de 259 pacientes con tumores de mama HER2+(7). El uso de la variable Ki67 antes del tratamiento como forma para predecir pCR tras el tratamiento con NAC se ha demostrado útil sobre todo en los tumores de mama ER-/HER2+ y TN, ya que son los que presentan mayor expresión de

rutas relacionadas con la proliferación. Un porcentaje alto de Ki67 (>50%) es un fuerte predictor de la respuesta al tratamiento con NAC (8) (9) (10) (11). Determinar los subtipos histológicos en base a principalmente tres marcadores como son el receptor de estrógenos (ER), el receptor de progesterona (PR) y la amplificación del gen HER2+, sirve para predecir también aquellos subgrupos de pacientes que tendrán un resultado concreto a largo plazo después del tratamiento. De esta manera, aquellas pacientes con tumores de mama triples negativos (TN) o con tumores HER2+, van a responder mejor al tratamiento con NAC que los subtipos positivos para los receptores hormonales (ER y PR) (93) (13) (14) (3) (4) (5).

En el escenario de los predictores de respuesta basados en el estudio de la biología molecular de los tumores, PAM50, como clasificador de subtipos basado en la expresión génica de 50 genes, se postula como el mejor predictor de respuesta al tratamiento con NAC ya que muchas son las publicaciones que avalan su capacidad predictiva. Estudios derivados de los ensayos clínicos NOAH, CherLOB, CALGB 40,601, PAMELA y BERENICE, entre otros, han demostrado la capacidad predictiva del clasificador PAM50 en tumores HER2+, todos ellos llegando a la conclusión de que aquellos tumores clasificados como HER2-*enriched* obtienen una alta tasa de pCR. De estos ensayos también se revela que los tumores con subtipo Luminal A determinados por PAM50 en tumores HER2+ por IHQ se asocia a los ratios más bajos de pCR (1) (15) (16) (17) (18) (19). Cuando todos los subtipos de cáncer de mama son incluidos, el subtipo Basal-Like se asocia a la ratio más alto de pCR. El subtipo Luminal A (asociado a su vez a perfiles con los mecanismos de proliferación menos expresados) es el que obtiene menor ratio de respuesta y por lo tanto es el subtipo de cáncer de mama definido por PAM50 que menos se beneficia del tratamiento con NAC (1) (2) (20) (15). La clasificación por IntClust de los tumores de mama, basado en las diferencias en la alteración en el número de copias que determinan diferentes subtipos y diferentes resultados clínicos, ha demostrado ser útil en la predicción de la respuesta a la NAC en una cohorte de 100 pacientes procedentes de los ensayos clínicos GEICAM/2006-03 y GEICAM/2006-14 (21). Por otra parte, en tumores HER2+, una firma de 41 genes, llamada TRAR, demostró capacidad predictiva de pCR en un estudio derivado del ensayo clínico NeoALTTO. TRAR obtuvo en su curva ROC valores

de AUC de 0.73 (61). Aun siendo algunos de estos marcadores muy prometedores, como es el caso de los subtipos moleculares de cáncer de mama definidos por PAM50, actualmente no hay un candidato claro que se use a nivel clínico para discernir pacientes respondedoras de no–respondedoras.

El objetivo de este estudio que reflejamos en el primer capítulo de la presente tesis es **conocer a nivel molecular las pacientes antes de ser tratadas con NAC y relacionar las diferencias encontradas con la respuesta al tratamiento** de forma retrospectiva. De esta forma, y aplicando nuestros conocimientos en el campo del *Machine Learning* e integrando las características clinicopatológicas, **definir una firma predictiva de pCR** en base a estas diferencias encontradas, que mejore el valor predictivo de los actuales marcadores. Debido al escaso material del que se dispone de las biopsias diagnósticas para investigación, una vez realizados todos los estudios clínico–patológicos necesarios para el adecuado tratamiento de la paciente, **solo se obtuvo material en un número suficiente de pacientes para llevar a cabo el estudio de la expresión génica y el diseño de una firma predictiva en base a estos resultados**. En cuanto al estudio de las alteraciones a nivel de DNA, como es el caso de la identificación de mutaciones, el tamaño muestral solo nos permite especificar una descriptiva de dichas mutaciones estratificada según respuesta.

4.2. POBLACIÓN DE ESTUDIO Y METODOLOGÍA

A continuación, se va a describir los pacientes y la metodología específica de este capítulo.

4.1.1 Pacientes y muestras

Se ha realizado un análisis retrospectivo con 111 pacientes diagnosticadas con cáncer de mama, tratadas con NAC antes de la cirugía. Las pacientes incluidas en este estudio se han seleccionado siguiendo criterios que describimos a continuación: mujeres adultas (mayores de 18 años) con cáncer de mama tratadas con NAC, es decir, Antraciclinas+/- Taxanos y terapia dirigida con Trastuzumab en aquellos tumores HER2+. Según el criterio Miller & Payne, se ha definido la pCR como la ausencia de carcinoma invasivo

tanto en la mama como en los ganglios linfáticos. Todos los análisis se han llevado a cabo en muestras de tejido previamente embebido en parafina desde el bloque de la biopsia diagnóstica antes del tratamiento Neoadyuvante. La respuesta al tratamiento (pCR vs no-pCR) se ha definido en la biopsia de la cirugía, tras el tratamiento Neoadyuvante. Debido a la limitación en el material de las biopsias diagnósticas, la cohorte de pacientes en el estudio del RNA y del DNA varía. Para el estudio de la expresión génica, la cohorte es de 84 pacientes (**RNA-cohorte**) mientras que para el estudio de las alteraciones a nivel de DNA es de 25 (**DNA-cohorte**). Debido a la limitación del número de pacientes, **el objetivo del capítulo 1 se va a llevar a cabo con la información de la expresión génica (RNA-cohorte)**, mientras que con la información de las alteraciones del DNA se hará una descriptiva de los resultados encontrados en la DNA-cohorte.

Como el objetivo de este capítulo es encontrar un biomarcador o firma predictiva de pCR en muestras de pacientes antes de ser tratadas con NAC, para darle más consistencia a nuestros resultados hemos validado nuestras firmas predictivas en una cohorte independiente. Por lo tanto, tenemos dos cohortes, nuestras pacientes donde hemos determinado los modelos predictivos (**Cohorte_Discovery**) y otra donde los hemos validado (**Cohorte_Validation**). Dentro de dichas cohortes hemos subdividido en grupos de pacientes según la sobreexpresión del gen HER2 y el estatus del receptor de estrógenos (ER). Por lo tanto, las cohortes han sido identificadas como **Allsubtypes**, que comprende las pacientes con todas las combinaciones posibles como son HER2-/ER-, HER2-/ER+ y HER2+ (tanto ER+ como ER-), **HER2-**, con solo aquellas pacientes con HER2- (independientemente del estatus de ER) y **ER+**, con aquellas pacientes ER+/HER2-. Cohortes HER2+ y otra ER-/HER2- no ha sido posible obtener debido al bajo número de pacientes con estas características.

Las cohortes de validación han sido extraídas de los estudios de Tabchy y colaboradores (87) (GEO DataSet: GSE20271) y Hatzis y colaboradores (88) (GEO DataSet: GSE25066), poseen el mismo esquema de tratamiento que la cohorte de nuestro estudio y era accesible la información tanto clínica-patológica como de los genes DE necesarios para el modelo predictivo.

Por lo que la distribución de las pacientes en el presente capítulo esta especificada en la Figura 11.

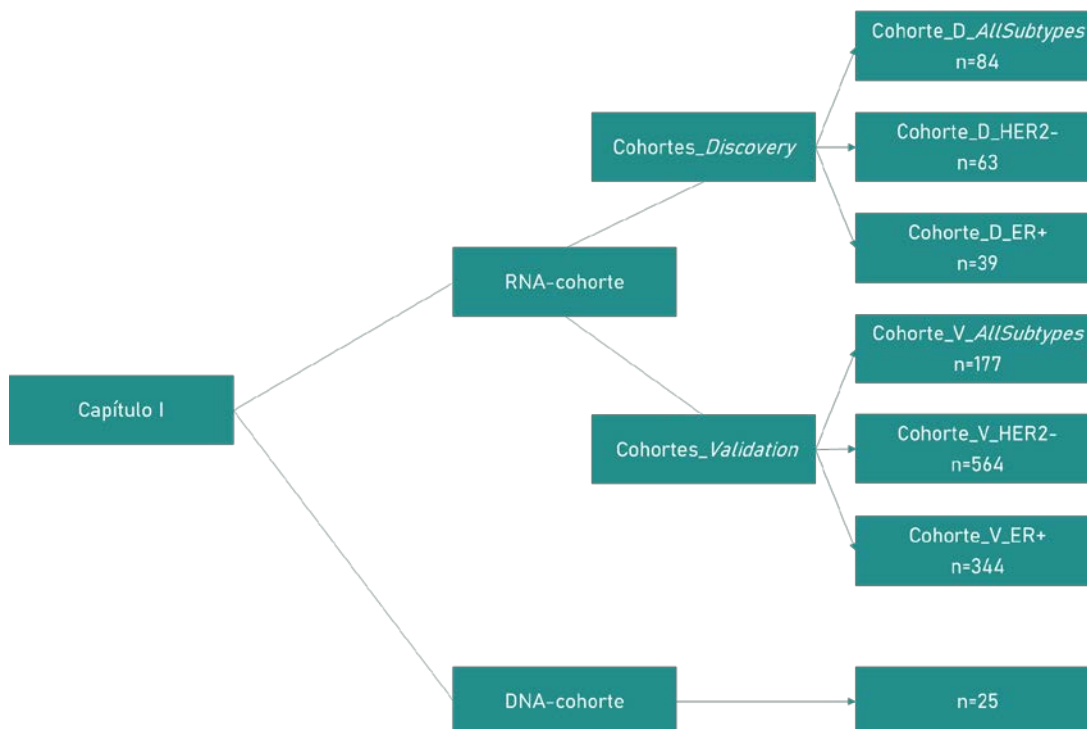


Figura 11. Distribución del número de pacientes de la RNA-cohorte según si han sido usados para el descubrimiento de los modelos predictivos de pCR tras el tratamiento con NAC o para la validación de ellos.

4.1.2 Inmunohistoquímica y determinación de la pCR

Desde los bloques de parafina de los pacientes seleccionados para el presente estudio se hizo cortes de 5µm para realizar las tinciones inmunohistoquímicas necesarias para determinar características como la expresión de; ER, PR, CK5/6, EGFR, Ki67 y HER2. Se usó una estación de tinción automatizada (Autostainer Plus, Dako), usando el EnVision FLEX System (Dako). Como cromógeno, usamos diamobenzidina y para reemplazar el anticuerpo (Ac) primario se usó un suero no-inmune como un control negativo, y un control positivo específico para cada anticuerpo.

Los anticuerpos que se usaron para la determinación de las proteínas específicas fueron: ER (Ac monoclonal de conejo, clona SP1, Master diagnóstica); PR (Ac monoclonal de conejo, clona Y85, Master diagnóstica); CK5/6 (Ac monoclonal de ratón, clona D5/16B4, Master Diagnóstica); EGFR (Ac monoclonal de conejo, clona EP38Y, Master Diagnóstica); Ki-67 (Ac monoclonal de conejo, clona SP6, Master Diagnóstica); y HER2 (Kit Herceptest,

Dako). La tinción fue evaluada por un patólogo. Para ER y PR el resultado positivo fue toda aquella tinción nuclear por encima del 1% (83) y para EGFR y CK5/6 cualquier tinción membranosa o citoplasmática. El Ki67 fue determinado cuantificando el porcentaje de células neoplásicas con tinción nuclear de Ki67. La sobreexpresión de HER2 fue evaluada de acuerdo a las guías internacionales (84) y categorizadas como negativas si el resultado de la inmunohistoquímica era 0 o +1; cuando el resultado era +2 se determinó la amplificación del gen por FISH o por CISH.

Para la determinación de la pCR, con una tinción de Hematoxilina y Eosina, el patólogo determinó un valor dentro de la clasificación de Miller and Payne, donde el 5 equivale a ausencia de células tumorales y del 4-1 equivale a diferentes grados de presencia de células tumorales. Es decir, todas aquellas pacientes con un valor 5 en la clasificación de Miller and Payne tanto en la muestra de la mama como en la muestra del ganglio linfático de la axila (ypT0/Tis, N0), se consideran buenas respondedoras al tratamiento con NAC y, por lo tanto, presentan pCR. Aquellas pacientes con valores entre 4 y 1 en dicha clasificación (en la muestra de la mama o en la axila) se consideran pacientes con tumores resistentes al tratamiento con NAC o, dicho de otra manera, pacientes con RD.

4.1.3 Expresión génica [RNA-cohorte]

4.1.3.1 Subtipos intrínsecos y el panel *PanCancer Pathways*

Los **subtipos moleculares** fueron determinados con el clasificador PAM50 de Nanostring, basado en la expresión de 50 genes y que determina 4 subtipos, Luminal A, Luminal B, HER2-enriched y Basal-Like (43). Por otra parte, para determinar la **expresión de los genes más importantes asociados a cáncer**, se realizó una medida de la expresión génica de 770 genes (730 genes diana + 40 genes *housekeeping*) con el panel *PanCancer Pathways*, mediante la tecnología nCounter®. Para determinar aquellos genes que se expresaban de forma diferencial entre pacientes que obtenían pCR tras el tratamiento versus aquellos que no, se realizó un **AED** con la herramienta nSolver™, donde el p-valor fue ajustado por el método Bejamini-Hochberg. Los genes DE con un p-valor < 0.05

fueron correlacionados entre ellos con el test de *Pearson* y el p-valor fue determinado por el *t-test*. Todo ello llevado a cabo en el entorno estadístico de R (versión 3.6).

Para analizar de una manera más global los cambios biológicos a nivel del conjunto de genes que conforman las funciones biológicas o módulos funcionales y no limitarnos a única entidad como es el gen, se ha realizado un **análisis de enriquecimiento de las rutas biológicas**. Basado en un análisis de componentes principales, la herramienta nSolver™ genera un *pathway-score* que resume la aportación de los genes de estudio a una ruta biológica concreta. Aplicándole la prueba *t* de *student* podemos determinar si existen diferencias significativas entre los pacientes con pCR y los que no.

4.1.3.2 Predicción de pCR

Con el objetivo de mejorar la capacidad predictiva de pCR de los biomarcadores ya conocidos, en las variables usadas para el desarrollo de modelos de predicción hemos incluido tanto las variables clínico-patológicas de estudio común, como otras variables también clínico-patológicas que nos parecían interesantes más los genes DE del panel *PanCancer Pathways*. Para asegurarnos de que la información de los genes DE aportaban valor predictivo, o incluso pudieran aportar más información por ellos mismos, dividimos los datos en 4 sets; **set de datos CLIN** (incluye solo la información clínico-patológica como los biomarcadores de pCR ya conocidos), **set de datos GENEX** (incluye solo los genes DE), **set de datos CLINGENEX** (incluye tanto la información clínico-patológica como los genes DE) y **PAM50** (incluye sólo la información de PAM50).

Todos los análisis que a continuación se especifican de los datos obtenidos y evaluados como predictores de pCR han sido desarrollados en el entorno estadístico R (versión 3.6).

Para determinar la relación de cada una de las variables independientes (características clínico-patológicas más genes DE) con la variable problema o dependiente (pCR), se realizó un **análisis univariante** basado en una RL usando como bondad de ajuste del modelo el LRT. Para estudiar si un conjunto de variables independientes puede afectar de manera combinada a la determinación de la variable dependiente hemos realizado

un **análisis multivariante** con doble validación cruzada. La validación interna con un modelo Lasso y una validación externa con LOOCV. Dicho **modelo predictivo multivariante** basado en LASSO, ha sido desarrollado para generar tres modelos predictivos para cada uno de los sets de datos (CLIN, GENEX, CLIN_GENEX) en las diferentes cohortes de descubrimiento (Cohorte_D_*Allsubtypes*, Cohorte_D_HER2- y Cohorte_D_ER+). Por lo tanto, los modelos predictivos van a ser llamados por el set de datos donde han sido definidos seguido de la cohorte donde ha sido probada, es decir, vamos a tener 9 modelos predictivos: **CLIN_*Allsubtypes*/HER2-/ER+**, **GENEX_*Allsubtypes*/HER2-/ER+**, y **CLINGENEX_*Allsubtypes*/HER2-/ER**. La importancia de cada variable dentro del modelo fue medida como el número de veces que fue seleccionada durante cada iteración. La firma del modelo la hemos definido como aquellas características que no contienen al 0 en el intervalo de confianza del 95%, es decir, que el 95% de las iteraciones el coeficiente de la variable es distinto a 0, o, lo que es lo mismo, que no se seleccionara el 0 más del 5% de las veces. Una vez definidos los modelos, la evaluación de ellos se realizó determinando el área bajo la curva ROC y AUC mediante la representación gráfica de la sensibilidad de nuestro modelo frente a la especificidad de este. Para hacer más robustos los resultados este proceso se repitió 1000 veces. El valor de dicho AUC se puede interpretar como la probabilidad de que un clasificador puntúe una instancia positiva aleatoriamente más alta que la negativa. Cuanto más lejana esté la línea de un modelo predictivo de la línea de estimación aleatoria de la curva ROC mayor poder predictivo tiene. Para evitar el **desequilibrio de las categorías dentro de las diferentes variables**, cuando el porcentaje de la clase minoritaria es inferior al 30% se ha usado el algoritmo SMOTE que sobremuestra dicha clase.

A parte de estos 3 modelos generados en base a los 3 sets de datos, hemos incluido un modelo que solo incluye a la variable PAM50 para ver el valor predictivo que tiene en nuestra Cohorte_*Discovery* por ella misma (**PAM50**).

Los resultados los vamos a representar mediante diferentes representaciones gráficas. Los **barcharts**, que nos presenta el número de veces que la variable es seleccionada por

LASSO durante las repeticiones de la validación externa por LOO. En los *crossbars* se muestra el coeficiente de cada variable de nuestro modelo. Como ha sido iterado 1000 veces, se ha identificado el cuartil 95% (IC del 95%). En los *violin plots* se puede observar gráficamente la distribución de los AUCs obtenidos durante las 1000 iteraciones. La forma de violín identifica la distribución que siguen dichos AUCs. Y los *violin ratios* donde se presenta la ratio de los AUCs de los modelos en cuestión, esto nos permite ver la relación entre dos modelos.

Una vez establecidos los modelos predictivos de pCR con los diferentes sets de datos y en los diferentes grupos de pacientes, se evaluó cada uno de ellos con la métrica AUC y con doble validación cruzada. Por último, para la **validación de nuestros modelos**, han sido ejecutados en cohortes externas que poseen el mismo esquema de tratamiento y poseen la información tanto clínica-patológica como de los genes DE necesarios para el modelo predictivo. Por lo tanto, podemos diferenciar entre la cohorte donde hemos realizado todos los experimentos y análisis relacionados con el diseño de los modelos predictivos (*Cohorte_Discovery*) y la cohorte externa donde hemos evaluado el poder predictivo de nuestros modelos (*Cohorte_Validation*). Como ya hemos explicado en el apartado de pacientes y muestras, la distribución de pacientes entre dichas cohortes se especifica en la Figura 10, y el flujo de trabajo para la determinación de los modelos predictivos, así como su validación, se especifican en la Figura 12.

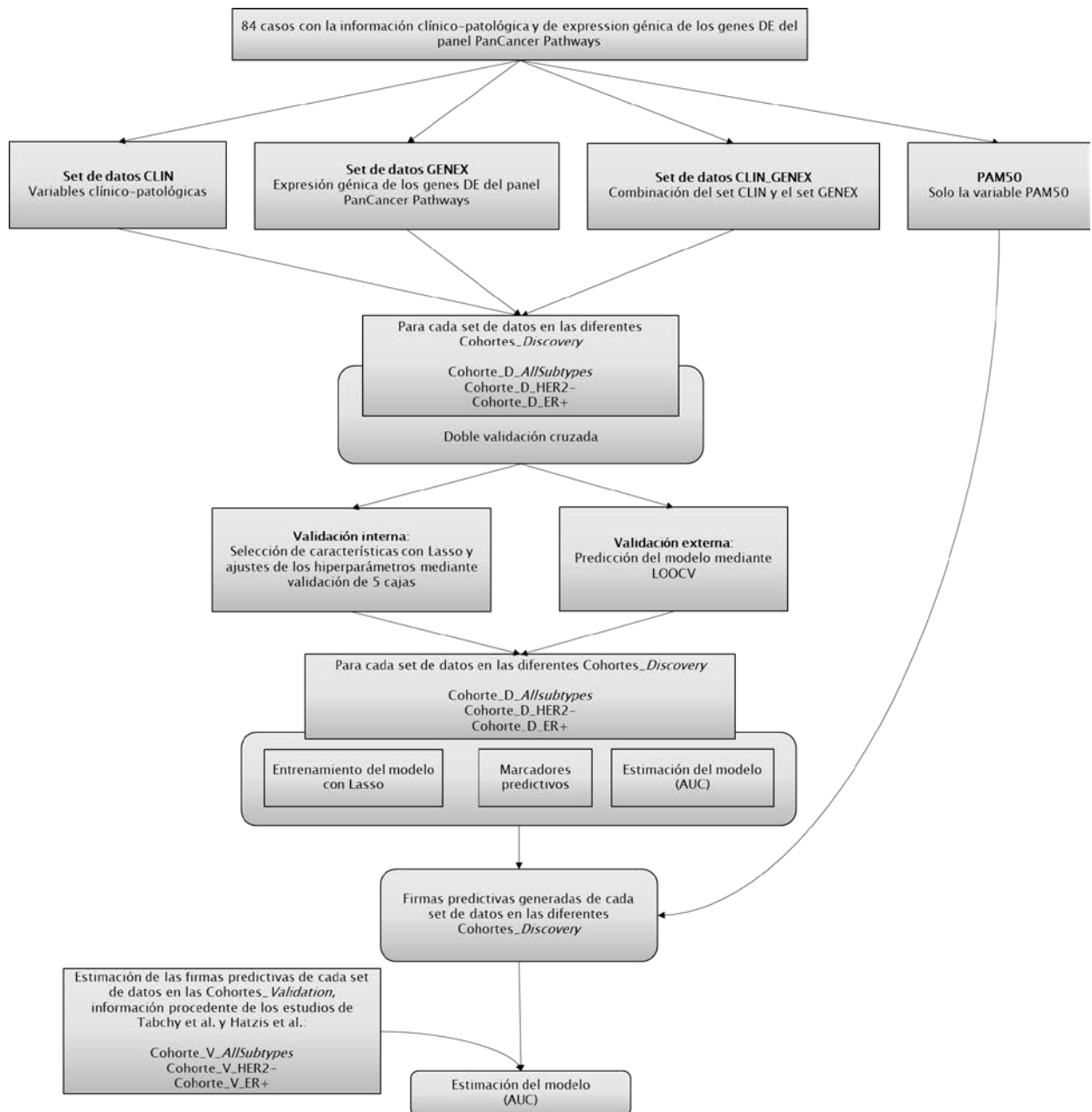


Figura 12. Flujo de trabajo para la determinación de los modelos predictivos. Con los sets de datos CLIN, GENEX y CLIN_GENEX se han determinado las variables más importantes para la determinación de pCR en las pacientes de las Cohortes_Discovery con una doble validación cruzada, una validación interna con Lasso y una validación externa por LOOCV. Paralelamente se ha incluido un modelo solo con la variable PAM50. De nuevo, Lasso ha sido el método usado para el entrenamiento de los modelos y la métrica para su estimación, AUC. Una vez las firmas han sido generadas, han sido probadas en las diferentes cohortes tanto en las Cohortes_Discovery como en las cohortes_Validation usando la métrica de AUC para su estimación. El modelo PAM50 solo ha sido probado en las Cohortes_Discovery.

4.1.4 Análisis de supervivencia

Usando el entorno estadístico R (versión 3.6) con el paquete *Survival* (<https://CRAN.R-project.org/package=survival>), estimamos la función de supervivencia, así como la curva de Kaplan-Meier para calcular la supervivencia global y los intervalos de confianza. Para los datos categóricos se ha usado el test log-rank para comparar las curvas de Kaplan-

Meier de diferentes grupos. En el caso de la expresión génica, se ha categorizado en base a los cuartiles en baja, intermedia y alta expresión. Además, también hemos llevado a cabo un modelo de regresión de Cox. La estimación de los parámetros se ha realizado por *partial likelihood*. El análisis univariante y multivariante ha sido realizado para las variables clínico-patológicas y la información de la expresión génica.

4.3. Mutaciones antes del tratamiento con NAC (DNA-cohorte)

El análisis bioinformático se realizó a través de la plataforma Genome one (Dreamgenics, S. L.), certificada con la marca IVD/CE de productos sanitarios para diagnóstico in vitro de acuerdo con la legislación vigente (Número de Licencia: 7157-PS). El procesamiento integra los siguientes pasos: Los archivos crudos FASTQ fueron evaluados usando los controles de calidad de FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y Trimmomatic (89) para la eliminación de bases, adaptadores y otras secuencias de baja calidad. Cada FASTQ se alineó frente a la versión GRCh37/hg19 del genoma humano de referencia mediante BWA-mem (90) y la generación de los archivos BAM ordenados se hizo con SAMtools (91). Al tratarse de un panel de amplicones, no se eliminaron lecturas duplicadas. SNVs e indels fueron identificadas mediante la combinación de VarScan 2 (92) y un algoritmo de llamada a variantes propio de Dreamgenics S.L. Las variantes se anotaron utilizando distintas bases de datos con información funcional (RefSeq, Pfam), poblacional (dbSNP, 1000 Genomes, ESP6500, ExAC, gnomAD), de predicción de impacto funcional *in silico* (dbNSFP, dbSNV), e información relacionada con el cáncer (COSMIC, ICGC).

En cuanto a la **priorización de variantes**, para la selección de variantes potencialmente somáticas en cada muestra se aplicaron los siguientes filtros generales:

- Variantes con presencia en COSMIC
- Predicción de efecto en proteína
- Frecuencia alélica poblacional inferior a 0.05
- Cobertura total en la posición afectada superior a 50X

- Frecuencia alélica de la variante en la muestra superior a 0.1

4.4. RESULTADOS

4.4.1. Expresión génica (RNA-Cohorte)

4.4.1.1. Características clínico-patológicas

A continuación, se detallan las descriptivas de las pacientes incluidas en este estudio. Se han dividido en las diferentes cohortes en las que se han definido las firmas predictivas para dar una idea al lector de cómo es la población de cada una de ellas y, a su vez, si son parte de la cohorte de descubrimiento (*Cohortes_Discovery*) o de la cohorte de validación (*Cohortes_Validation*).

En las *Cohortes_Discovery* (Tabla 5) la mayoría de las pacientes son pre-menopáusicas (70-79% entre las cohortes Allsubtypes, HER2- y ER+), tienen un tamaño tumoral prequirúrgico medio en torno a 3.7 cm, el 58.3% de las pacientes tienen afectación ganglionar, el 56% de los tumores tienen un grado tumoral histológico entre 1 y 2; cabe destacar que el 74% las pacientes de la cohorte ER+ tienen el grado histológico entre 1 y 2. En el caso de la cohorte que contempla todos las pacientes (*Cohorte_D_Allsubtypes*) el 60% son ER+, el 25% son HER2+ y el 63% tiene menos del 50% de expresión en el marcador de proliferación Ki67. En cuanto a los subtipos definidos por IHQ, el 40% son Luminales, el 25% HER2+ y el 27% son TN. Los subtipos moleculares definidos por PAM50 están distribuidos de la siguiente manera, el 10.7% son Luminales A, el 33.33 % Luminal B, el 26.19% son HER2-enriched y el 29% son Basal-Like. De esta cohorte, el 33.33% han obtenido pCR mientras que el 66.67% no.

En la cohorte específica de pacientes con el marcador HER2 negativo (*Cohorte_D_HER2-*), el 39% tienen ER+, ninguna sobreexpresa el HER2 (es lo que define al grupo), y el 57% de las pacientes tienen una expresión de Ki67 menor del 50%. Un 63.5% son Luminales y el otro 36.5% son TN según los subtipos definidos por IHQ. Encontramos la misma proporción de pacientes Luminal B y de Basal-Like (39.7%), mientras que el subtipo de PAM50 menos presente es el HER2-enriched, solo un 8% de las pacientes. El 25.4% de las pacientes obtienen pCR.

En la cohorte específica de pacientes con HER2- y ER+ (Cohorte_D_ER+), todas las pacientes expresan el marcador ER y ninguna sobreexpresa el HER2. El 70% de las pacientes expresan menos del 50% del marcador Ki67, y el 97.4% de las pacientes son subtipos Luminal definido por IHQ. El 65% son Luminal B y el 15.7% Luminal A definido por PAM50. El 79.5% de las pacientes no obtienen pCR en esta cohorte. En la tabla 6 se han descrito las diferencias clínico-patológicas entre las pacientes que obtienen pCR de las que no en la cohorte_D_AllSubtypes.

Tabla 5. Características clínico-patológicas de las pacientes de las cohortes donde los modelos han sido definidos, Cohortes_Discovery

Variables Clínico-patológicas	Cohorte_D_Allsubtypes		Cohorte_D_HER2-		Cohorte_D_ER+	
	n	%	n	%	n	%
Todos los casos	84	100	63	100	39	100
Estatus Menopáusico						
Pre-menopáusico	59	70.23	45	71.42	31	79.49
Post-menopáusico	25	29.77	18	28.58	8	20.51
Tamaño tumoral (media)	3.7	-	3.9	-	3.7	-
Afectación ganglionar clínica						
Positiva	49	58.33	35	55.56	21	53.85
Negativa	35	41.67	28	44.44	18	46.15
Grado tumoral histológico						
1-2	47	55.95	34	53.97	29	74.36
3	37	44.05	29	46.03	10	25.64
Expresión del ER						
Positivo	50	59.52	39	61.9	39	100
Negativo	34	40.48	24	38.1	0	0
Expresión del PR						
Positivo	42	50	33	52.38	32	82.05
Negativo	42	50	30	47.62	7	17.95
Sobrexpresión del HER2						
Positivo	21	25	0	0	0	0
Negativo	63	75	63	100	39	100
Ki67						
< 50%	53	63.1	36	57.14	27	69.23
> 50%	31	36.9	27	42.85	12	30.77
Subtipos IHQ						
Luminales (A+B)	40	47.62	40	63.49	38	97.44
HER2+	21	25	0	0	0	0
Triples Negativos	23	27.38	23	36.51	1	2.56
Subtipos moleculares por PAM50						
Luminal A	9	10.71	8	12.70	6	15.79
Luminal B	28	33.33	25	39.68	25	65.79
HER2-enriched	22	26.19	5	7.94	4	10.53
Basal-Like	25	29.77	25	39.68	3	7.89
Respuesta completa patológica (pCR)						
Si	28	33.33	16	25.4	8	20.51
No	56	66.67	47	74.6	31	79.49

Tabla 6. Características clínico-patológicas de las pacientes incluidas en la Cohorte_D_Allsubtypes según si han respondido o no al tratamiento con NAC.

Variables clínico-patológicas	<i>pCR</i>		<i>no-pCR</i>	
	n	%	n	%
Estatus menopáusico				
Premenopáusica	17	61	42	75
Postmenopáusica	11	39	14	25
Tamaño tumoral				
Menor de 2 cm	1	4	10	18
Entre 2 y 5 cm	23	82	33	59
Más de 5 cm	4	14	13	23
Afectación ganglionar clínica				
Positivo	15	54	34	61
Negativo	13	46	22	39
Grado tumoral histológico				
1	1	4	7	13
2	13	46	26	46
3	14	50	23	41
Estatus ER				
Positivo	15	54	35	63
Negativo	13	46	21	38
Estatus PR				
Positivo	12	43	30	54
Negativo	16	57	26	46
Estatus HER2				
Positivo	12	43	9	16
Negativo	16	57	47	84
% Ki67				
Menor que 50	16	57	37	66
Más de 50	12	43	19	34
Subtipos IHQ				
Luminal A	0	0	4	7
Luminal B	8	29	28	50
HER2	12	43	9	16
Triple Negativo	8	29	15	27
Subtipos PAM50				
Luminal A	3	11	6	11
Luminal B	5	18	23	41
HER2-enriched	11	39	11	20
Basal-Like	9	32	16	29

De las **Cohortes_Validation** (Tabla 7) hay muchos datos de las variables clinicopatológicas que no disponemos debido a que en los estudios de origen no están recogidas, como es el caso de los subtipos definidos por IHQ, los subtipos moleculares

definidos por PAM50 o el % de expresión de Ki67. En términos generales, la media del tamaño tumoral prequirúrgico está entre 2.7 y 3.17 cm, más del 60% de las pacientes tienen afectación ganglionar, y el grado tumoral histológico oscila entre 1, 2 y 3; cabe destacar, como en la cohorte *Discovery*, que el menor porcentaje de las pacientes con grado histológico 3 están en la cohorte ER+. En el caso de la cohorte que contempla todas las pacientes (Cohorte_V_*Allsubtypes*) el 54% son ER+, el 14.5% son HER2. De esta cohorte solo el 14.69% han obtenido pCR. En la cohorte específica de pacientes con el marcador HER2 negativo (Cohorte_V_HER2-), el 61% tienen ER+ y aproximadamente el 17% de las pacientes obtienen pCR. En la cohorte específica de pacientes con HER2- y ER+ (Cohorte_V_ER+), todas las pacientes expresan el marcador ER y ninguna sobreexpresa el HER2. Más del 91% de las pacientes no obtienen pCR en esta cohorte.

Podemos destacar como diferencias entre ambas cohortes (*Discovery vs Validation*), que el tamaño tumoral medio entre una y otra varía, conteniendo tumores más grandes en la Cohorte *Discovery* (3.7 vs 2.9 cm, respectivamente). La afectación ganglionar es levemente mayor en la Cohorte *Validation* y la representación de tumores positivos para ER es similar entre ambas cohortes, mientras que la sobreexpresión de HER2 es mayor en la Cohorte *Discovery*.

Tabla 7. Características clínico-patológicas de los pacientes incluidos para la validación de los modelos predictivos, Cohortes_Validation

Variables Clínico-patológicas	Cohorte_V_Allsubtypes		Cohorte_V_HER2-		Cohorte_V_ER+	
	n	%	n	%	n	%
Todos los casos	177	100	564	100	344	100
Tamaño tumoral (media)	2.7	-	3.17	-	3.14	-
0	2	1.12	2	0.35	2	0.58
1	11	6.22	12	2.13	10	2.9
2	75	42.38	85	15.07	52	15.12
3	37	20.9	254	45.04	153	44.48
4	51	28.82	211	37.41	127	36.92
NA	1	0.56				
Afectación ganglionar clínica						
Positivo	118	66.66	377	66.85	219	63.66
Negativo	59	33.34	187	33.15	125	36.34
Grado tumoral histológico						
1	15	8.47	42	7.45	38	11.05
2	60	33.9	220	39	186	54.07
3	72	40.68	302	53.55	120	34.88
NA	30	16.95				
Expresión del ER						
Positivo	97	54.8	344	61	344	100
Negativo	80	45.2	220	39	0	0
Expresión del PR						
Positivo	82	46.32	282	50	262	76.16
Negativo	95	53.68	282	50	82	23.84
Sobreexpresión del HER2						
Positivo	26	14.68	0	0	0	0
Negativo	151	85.32	564	100	344	100
Respuesta completa patológica (pCR)						
Si	26	14.69	97	17.2	28	8.14
No	151	85.31	467	82.8	316	91.86

4.1.4.1 Asociación entre variables clínico-patológicas

En lugar de un análisis de correlación, que únicamente abarca aquellas correlaciones con relación lineal, se ha realizado un análisis de asociación entre variables para poder identificar todas las asociaciones, tanto lineales como no lineales. En la tabla se especifican aquellas asociaciones con significación estadística. En nuestra cohorte_Discovery hemos encontrado que las asociaciones más fuertes son entre la

expresión de ER y de PR ($p < 1.02e-13$), entre la expresión de ER y los subtipos de PAM50 ($p < 3.68e-8$) y entre los subtipos de PAM50 y los subtipos IHQ ($p < 1e-7$) (Tabla 8).

Tabla 8. Asociaciones entre las variables incluidas en este estudio. Solo se muestran aquellas con significación estadística

Variable clínico-patológica 1	Variable clínico-patológica 2	Test estadístico
Grado tumoral histológico	Expresión de ER	0.0007548
Grado tumoral histológico	Expresión de PR	0.008355
Grado tumoral histológico	Ki67	5.61e-05
Grado tumoral histológico	Subtipos IHQ	3.32e-05
Grado tumoral histológico	Subtipos PAM50	0.002618
Expresión de ER	Expresión de PR	1.021e-13
Expresión de ER	Subtipos IHQ	1e-07
Expresión de ER	Subtipos PAM50	3.682e-08
Expresión de PR	Subtipos IHQ	1e-07
Expresión de PR	Subtipos PAM50	1.105e-05
Sobreexpresión de HER2	Subtipos IHQ	1e-07
Sobreexpresión de HER2	Subtipos PAM50	0.01054
Sobreexpresión de HER2	PCR	0.01616
Ki67	Subtipos IHQ	0.003621
Ki67	Subtipos PAM50	0.005059
Subtipos IHQ	Subtipos PAM50	1e-07
Subtipos IHQ	PCR	0.01371

4.4.1.2. Panel *PanCancer pathways* en la cohorte *Discovery*. Análisis de expresión diferencial (AED), correlación entre genes diferencialmente expresados y enriquecimiento en rutas biológicas.

Con el objeto de conocer cómo se comportan los genes, seleccionados en base a su asociación con cáncer, entre las pacientes que obtienen pCR y las que no (en la Cohorte_D_*Allsubtypes*), y, en base a esa información, encontrar una firma predictiva de la respuesta al tratamiento, se ha realizado un AED de los 730 genes incluidos en el panel *PanCancer Pathways* de Nanostring.

De las 84 pacientes incluidas en este estudio se pudo obtener **RNA de calidad** (según ratios 260/280 y 260/230) con una concentración óptima para llevar a cabo la medición de la expresión génica.

Los **genes DE** (Tabla 9) entre las pacientes que obtienen pCR y las que no, con p-valor por debajo de 0.05 son: *CDC6* ($p=0.00016$), *LEFTY1* ($p=0.00024$), *FGFR4* ($p=0.0018$),

MYCN (p=0.0092), *IDH2* (p=0.0092), *STK11* (p=0.034), *FGF10* (p=0.041), *RASGRF1* (p=0.045), *WNT3* (p=0.045) y *GNG4* (p=0.045). Los datos de log₂ *Fold Change* (FD) indica que los genes *CDC6*, *FGFR4*, *MYCN*, *IDH2*, *RASGRF1* y *GNG4* se encuentran más expresados en aquellas pacientes que consiguen pCR tras el tratamiento con NAC que en aquellas que no. El opuesto es el caso de los genes *LEFTY1*, *STK11*, *FGF10*, *WNT3*, que se encuentran más expresados en las pacientes con tumores resistentes al tratamiento, es decir con RD. Los mayores cambios encontrados entre pCR y no pCR se observan en los genes *CDC6*, *LEFTY1*, *FGFR4*, *MYCN* y *GNG4*, con un log₂ FC superior a 1.5 en valor absoluto (Tabla 9). En la figura 13 se representan el número de *counts* normalizados determinados en la medición de la expresión génica de cada uno de los genes DE entre las pacientes que obtienen pCR y las que no. En el *volcano plot*, podemos observar aquellos genes DE con significancia estadística (p<0.05) y el log₂ FC de cada uno de ellos (Figura 14).

Tabla 9. Genes DE entre las pacientes que obtienen pCR tras el tratamiento con NAC y las que no. Solo se han seleccionado aquellos genes con un p-valor ajustado por debajo de 0.05. Se incluye una barra de intensidad según el Log₂ FC que se obtiene de cada gen. En escala de rojo aquellos genes con un Log₂ FC superior a 0, es decir que se expresan más en aquellos pacientes que obtienen pCR y en escala de azul aquellos que se expresan menos.

Genes	Log ₂ <i>Fold Change</i>	p-valor ajustado por Bejamini-Hochberg	Rutas biológicas
<i>FGFR4</i>	1.81	0.00185	MAPK, PI3K, RAS
<i>CDC6</i>	1.68	0.000167	<i>Cell Cycle - Apoptosis</i>
<i>GNG4</i>	1.59	0.045	PI3K, RAS
<i>MYCN</i>	1.58	0.00922	<i>Transcriptional Misregulation</i>
<i>RASGRF1</i>	1.39	0.045	MAPK, RAS
<i>IDH2</i>	0.882	0.00922	<i>Driver Gene</i>
<i>STK11</i>	-0.359	0.0349	<i>Driver Gene</i> , PI3K
<i>WNT3</i>	-0.891	0.045	<i>Hedgehog</i> , Wnt
<i>FGF10</i>	-1.41	0.0418	MAPK, PI3K, RAS
<i>LEFTY1</i>	-1.51	0.00024	TGF-beta

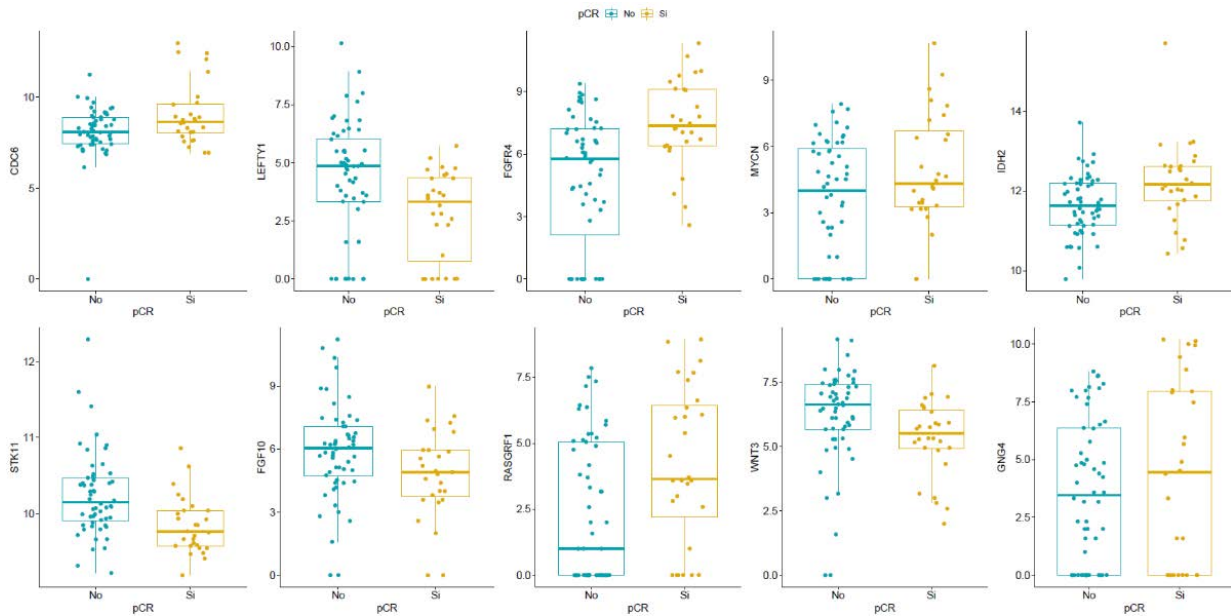


Figura 13. Diagramas de cajas. Se representan cada uno de los genes DE con un diagrama de cajas donde se muestra el número de *counts* normalizados procedentes del análisis de los resultados obtenidos (eje de ordenadas) frente a la situación de estudio, pCR sí o no (eje de abscisas).

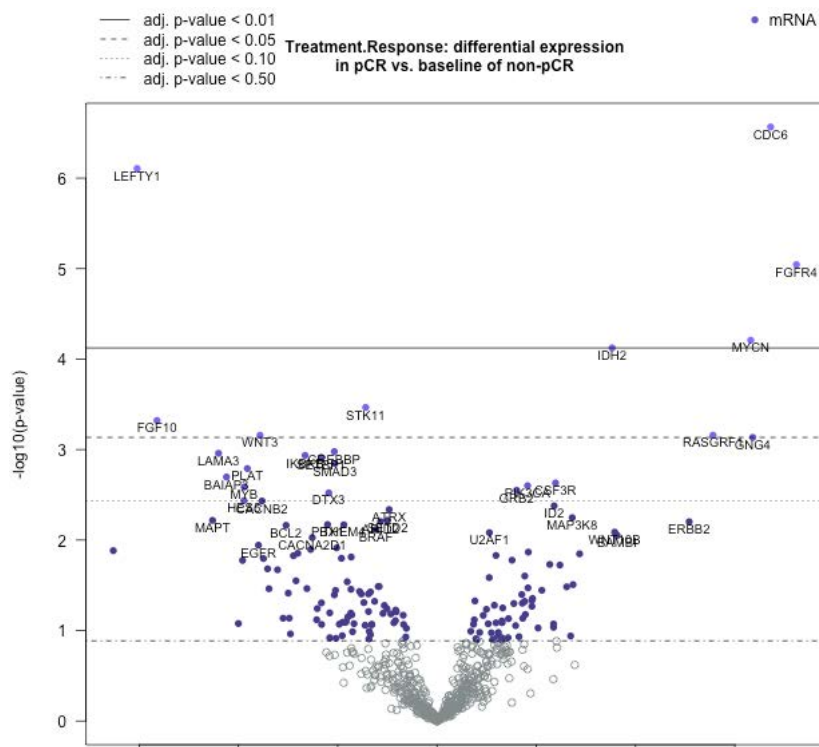


Figura 14. En el *volcano plot* se representan todos los genes según el p-valor ajustado obtenido en el AED (eje de ordenadas) y el valor del Log2 FC (eje de abscisas). Con las líneas punteadas se representan diferentes niveles de p-valor y la zona del volcano donde se posicione el gen corresponderá con un p-valor concreto y un valor de FC determinado. Aquellos genes en la zona izquierda desde el valor 0 del eje de abscisas, serán aquellos genes con menor expresión en las pacientes que obtienen pCR, en la parte derecha estarán los genes que se expresen al alza.

A continuación, vamos a detallar la función de cada uno de los genes DE. La mayoría de la información ha sido extraída de la *website Genecard* (www.genecard.com) y se ha especificado el año en el que se incluyó la información del gen en RefSeq.

EL gen *CDC6* es un regulador esencial en la activación y mantenimiento del ciclo celular (RefSeq, 2008). En tumores de mama ER-, la sobreexpresión de *CDC6* se ha visto relacionada con mal pronóstico a través de la sobreexpresión, a su vez, de MCM2-7(94). La alta expresión de *CDC6* está asociada a un aumento en la proliferación celular (95) por lo que cabe esperar una respuesta satisfactoria al tratamiento con NAC en aquellos pacientes que tengan sobreexpresado este gen, como es el caso de nuestros resultados.

EL gen *LEFTY1* ha sido relacionado con fenotipos *Stemness* (96), está involucrado en la activación de la proliferación de las células progenitoras epiteliales tanto tumorales como sanas (23). La expresión al alza de este gen está relacionada con la capacidad proliferativa a largo plazo de las células progenitoras y esto a su vez se ha demostrado que se asocia a resistencia al tratamiento tanto con radiación como con quimioterapia (97). En nuestros resultados, los pacientes que obtienen pCR, es decir, que son sensibles al tratamiento con NAC el cual se basa en el uso de agentes quimioterápicos, tienen una disminución en la expresión de este gen con respecto a aquellos pacientes con tumores resistentes. Estos resultados tienen coherencia con lo encontrado en la literatura. *Wnt3*, que está también involucrado en procesos que tienen lugar durante la embriogénesis, es un oncogén ampliamente estudiado y relacionado con el cáncer de mama. Se relaciona con marcadores de transición epitelio-mesénquima (EMT, *epithelial-mesenchymal transition*) en tumores de mama ER- (98). La expresión de *Wnt3* es más elevada en el desarrollo y durante la morfogénesis embrionaria mamaria; por otra parte, su importancia en la activación de las *stemcells* ya ha sido validada (99). Al igual que lo esperado con la expresión del gen *LEFTY1*, el hecho de que en nuestra cohorte el gen *Wnt3* se encuentre más inhibido en aquellos tumores sensibles al tratamiento, podría deberse a la menor presencia de *cancer stem cells* en las muestras sensibles a la NAC. En el caso de *FGF10* y *FGFR4*, ambos están relacionados con el crecimiento fibroblástico; *FGF10* tiene capacidad mitogénica y está involucrado en el desarrollo, crecimiento tumoral, invasión y supervivencia celular (RefSeq, 2008); además está relacionado con

EMT en el proceso de iniciación metastásico en cáncer de mama (100). Estas herramientas de supervivencia celular del gen *FGF10*, son coherentes en nuestros resultados con su expresión a la baja en los tumores que obtienen pCR. *FGFR4* es una tirosin-kinasa que actúa como receptor citoplasmático para los factores de crecimiento fibroblástico (RefSeq, 2017). La sobreexpresión de este gen ha sido previamente definida y asociada al subtipo HER2-enriched. Teniendo en cuenta que dicho subtipo obtiene la mejor tasa de pCR tras el tratamiento con NAC (1), cabe esperar que la alta expresión de *FGFR4* se asocie con tumores que responden al tratamiento en nuestra cohorte, tal y como nuestros resultados muestran. *MYCN* es un proto-oncogén, factor de transcripción de la familia MYC. Un estudio reciente en una cohorte de tumores TN ha demostrado que dicho gen está sobreexpresado significativamente en tumores resistentes a la NAC pero no en tumores que no han sido tratados, en ambos casos después de la cirugía (101). Según nuestros resultados una expresión al alza de este gen en el tumor primario se asocia a respuesta al tratamiento con NAC satisfactoria. Dado que nuestra cohorte no es exclusivamente de tumores TN, y que los estudios más actuales sobre este gen no son como biomarcador predictivo sino relacionado con un estatus concreto tras el tratamiento, no encontramos resultados contradictorios. *IDH2* tiene un rol importante en la producción de energía celular; mutaciones en este gen están asociadas a cánceres como el de colon o de cerebro (RefSeq, 2014). La sobreexpresión de este gen en líneas celulares de cáncer de pulmón se ha visto asociada a un aumento del crecimiento celular mediante la bajada de α KG y ROS, y la inducción de HIF1 α (efecto Warburg) (102). Dado que el tratamiento con NAC es más efectivo cuando hay una activación de la proliferación, cabe pensar que, aunque la patología no sea la misma, el efecto en la maquinaria celular sea similar y eso coincide con los resultados obtenidos. *STK11* es un miembro de la familia Serina/Treonina kinasa. Está implicado en procesos como el metabolismo celular, apoptosis, respuestas al daño celular o freno del ciclo celular (103). Actúa como un supresor de tumores a través del control de la familia AMPK (RefSeq, 2008). Su estrecha relación con el control y freno del ciclo celular concuerda con que esté menos expresada en pacientes que responden al tratamiento y que podrían tener más activada la proliferación. Por último, *RASGRF1*,

responsable de la disociación de GDP de las proteínas RAS (RefSeq, 2009), también ha sido postulado como un marcador epigenético en ctDNA para la detección de cáncer de mama temprano (104). Por último, **GNG4** es la subunidad gamma 4 de la proteína G (UniProtKB P50150). A nivel de respuesta ante el tratamiento citotóxico, diferencias en la expresión génica de estos genes no han sido relacionadas previamente con ella.

En el *circus plot* (Figura 15) ha sido representado los genes DE y las rutas biológicas a los que pertenecen esos genes (los 13 vías canónicas que tanto el panel *PanCancer Pathways* como la herramienta nSolver™ tienen en cuenta). Representados de esta manera podemos observar como las rutas biológicas de la PIK3 y de RAS están más representados en nuestros datos, ya que son los que comparten más genes DE. Esto solo nos da una idea, ya que para saber si existen enriquecimiento de vías biológicas entre las condiciones pCR y no pCR es necesario usar la información de todos los genes analizados y no solo los genes DE, además de hacer otros tests estadísticos como la *t* de *Student*. Aunque lo vamos a especificar más adelante, ya podemos adelantar que la vía *Transcriptional Regulation* es uno de los que se relacionan con la pCR de manera estadísticamente significativa.

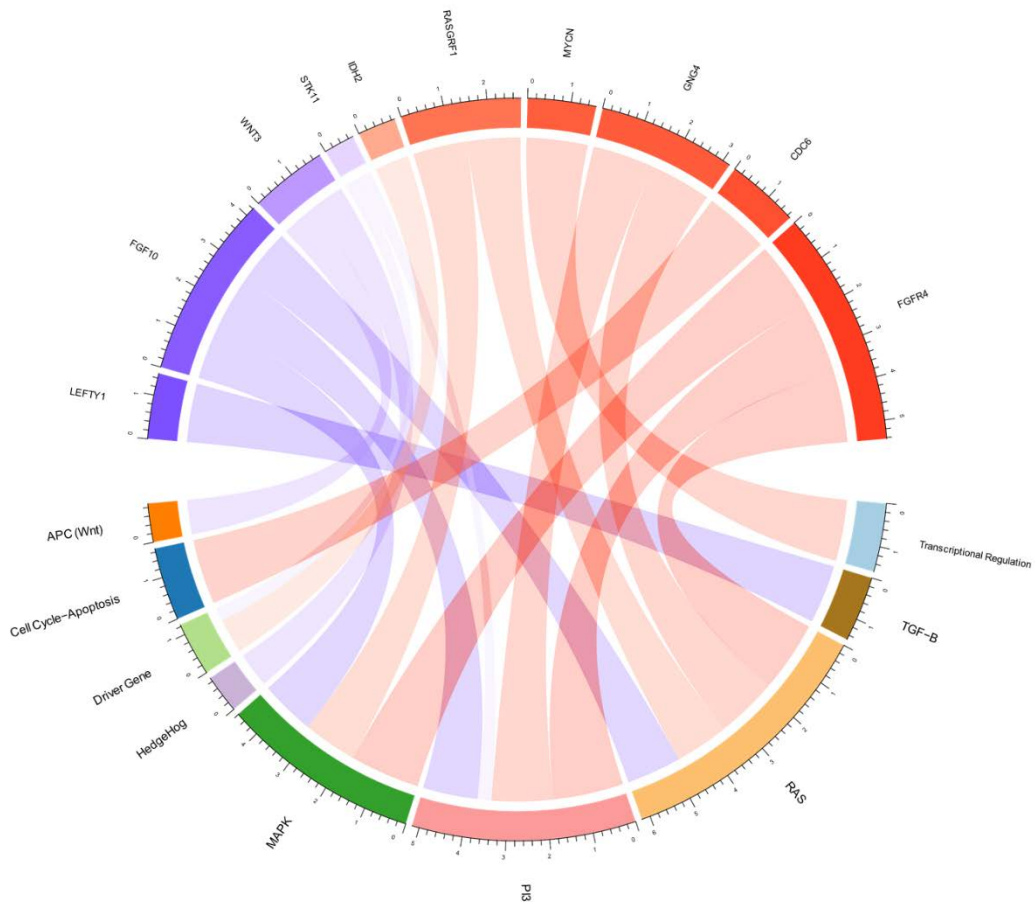


Figura 15. *Circos plot.* Se representa los 10 genes DE enlazados a las rutas que pertenecen con colores diferentes. Los genes, a su vez, se le asigna un color que va desde el rojo hasta el azul según el Log₂ FC que presenten con respecto a la pCR.

En la Figura 16 y en la Tabla 10 se muestran los resultados del análisis de correlación entre los genes DE. En la figura 15 vemos cómo se van moviendo las relaciones entre los genes desde -1 (correlación negativa perfecta, cuando uno aumenta el otro disminuye) hasta 1 (correlación positiva perfecta, cuando uno aumenta el otro también). El signo de la correlación entre genes diferencialmente expresados en la misma dirección (por ejemplo, sobreexpresados), es positivo en la mayoría de los casos, mientras que es negativo en aquellos genes con tendencias opuestas con respecto a la expresión diferencial.

Tabla 10. Esta tabla dividida en dos representa el valor r de la correlación de *Pearson* de cada uno de los pares de genes y el p -valor del t -test asociado a dicha correlación.

Correlación de Pearson	CDC6	LEFTY1	FGFR4	MYCN	IDH2	STK11	FGF10	RASGRF1	WNT3	GNG4
CDC6	1	-0.175	0.298	0.129	0.229	-0.352	-0.076	0.239	0.244	0.178
LEFTY1	-0.17	1	-0.101	-0.101	-0.209	0.360	0.153	-0.357	0.183	-0.12
FGFR4	0.298	-0.101	1	0.390	0.242	-0.171	-0.184	0.281	-0.219	0.069
MYCN	0.129	-0.101	0.390	1	0.004	-0.324	-0.010	0.301	-0.014	0.371
IDH2	0.229	-0.209	0.242	0.004	1	-0.290	-0.201	0.235	-0.271	0.065
STK11	-0.35	0.360	-0.171	-0.324	-0.290	1	0.057	-0.255	-0.037	-0.30
FGF10	-0.07	0.153	-0.184	-0.010	-0.201	0.057	1	-0.252	0.333	-0.08
RASGRF1	0.239	-0.357	0.281	0.301	0.235	-0.255	-0.252	1	-0.233	0.255
WNT3	0.244	0.183	-0.219	-0.014	-0.271	-0.037	0.333	-0.233	1	0.042
GNG4	0.178	-0.120	0.069	0.371	0.065	-0.306	-0.089	0.255	0.042	1
p-valor del t-test	CDC6	LEFTY1	FGFR4	MYCN	IDH2	STK11	FGF10	RASGRF1	WNT3	GNG4
CDC6		0.1120	0.0059	0.2421	0.0360	0.0010	0.4930	0.0288	0.0251	0.105
LEFTY1	0.112		0.3593	0.3622	0.0570	0.0008	0.1651	0.0009	0.0950	0.277
FGFR4	0.005	0.3593		0.0002	0.0266	0.1191	0.0932	0.0097	0.0454	0.531
MYCN	0.242	0.3622	0.0002		0.9733	0.0026	0.9314	0.0053	0.9004	0.000
IDH2	0.036	0.0570	0.0266	0.9733		0.0074	0.0664	0.0311	0.0127	0.554
STK11	0.001	0.0008	0.1191	0.0026	0.0074		0.6043	0.0194	0.7359	0.046
FGF10	0.493	0.1651	0.0932	0.9314	0.0664	0.6043		0.0209	0.0019	0.419
RASGRF1	0.028	0.0009	0.0097	0.0053	0.0311	0.0194	0.0209		0.0330	0.019
WNT3	0.025	0.0950	0.0454	0.9004	0.0127	0.7359	0.0019	0.0330		0.706
GNG4	0.105	0.2776	0.5316	0.0005	0.5542	0.0046	0.4196	0.0192	0.7062	

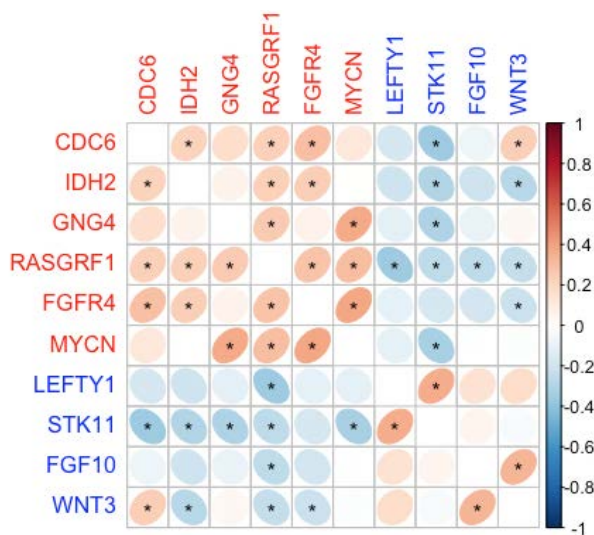


Figura 16. En la matriz se muestran las correlaciones entre los 10 genes DE. El color de los nombres de los genes depende del Log2 FC obtenido en el AED (el color rojo para los genes expresados al alza en las pacientes con pCR y en azul aquellos expresados a la baja). En la barra lateral vemos una gradación de color desde rojo hasta azul que corresponde al valor r de la correlación de Pearson. Desde el rojo más intenso (correlación positiva perfecta, $r=1$) hasta el azul más oscuro (correlación negativa perfecta, $r=-1$). Esto mismo se ha representado en cada una de las interacciones dentro de la matriz. Aquellas interacciones que estadísticamente significativas se han identifica con un asterisco.

Aunque en los resultados del AED ya veíamos una tendencia en cuanto a las rutas biológicas más representados en el set de genes DE, para determinar el **enriquecimiento en rutas biológicas** entre los grupos de pacientes pCR y no-pCR es necesario tener la información del mayor número de genes implicados en dichas vías, por ello usamos el *pathway-score* (unifica los datos de cada gen que forma parte de esa ruta biológica en un único *score*) y la *t* de *student* para ver si existen diferencias significativas. De las 13 vías o rutas que contemplamos, en los resultados del *t* de *student* se muestra que existen 3 vías que se encuentran más activadas en una situación con respecto a otra. Las vías *Chromatin modification* ($p=0.01$), *Transcriptional Regulation* ($p=0.032$) y *JAK-STAT* ($p=0.048$) se encuentran más activados en aquellas pacientes que consiguen pCR tras el tratamiento con NAC (Figura 17).

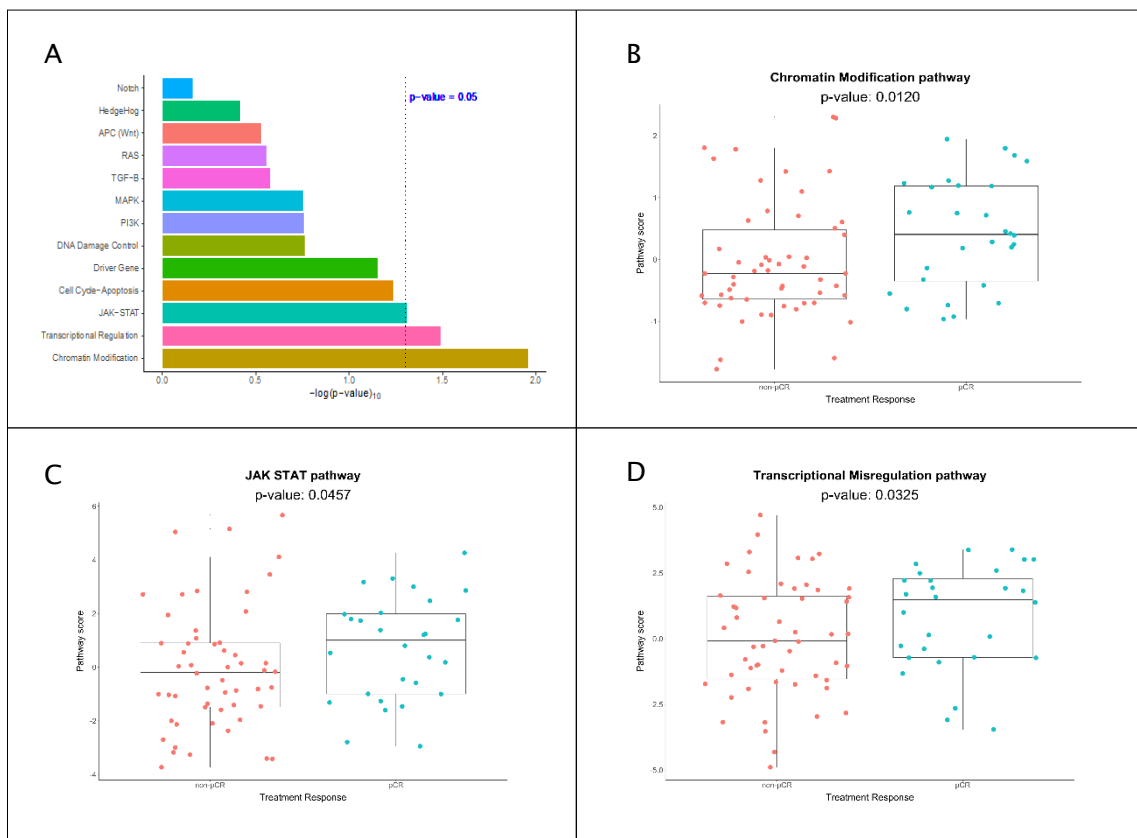


Figura 17. A. Diagrama de barras. Representación de todas las vías o rutas biológicas que incluimos en el análisis funcional. La línea punteada determina el p-valor por debajo de 0.05. B, C y D. Diagramas de cajas. Se representan las tres rutas estadísticamente significativas según su comportamiento entre las pacientes que obtienen pCR y las que no, *Chromatin modification*, JAK-STAT y *Transcriptional misregulation pathways*.

4.4.1.3. Predicción de pCR en Cohorte_*Discovery*

Tal y como hemos especificado con anterioridad, actualmente hay variables clínico-patológicas ya conocidas que predicen pCR, sin embargo, ninguno de ellos se posiciona como un claro marcador de la respuesta a la NAC en la comunidad científica. Con el objetivo de mejorar la capacidad predictiva de pCR de los biomarcadores ya conocidos, en las variables usadas para el diseño de modelos predictivos hemos incluido tanto las variables clínico-patológicas predictoras de pCR ya conocidas, incluyendo la variable de subtipos moleculares definidos por PAM50, como otras variables también clínico-patológicas que nos parecían interesantes y los genes diferencialmente expresados del panel *PanCancer Pathways* entre aquellos pacientes con pCR y no-pCR.

Según el **análisis univariante**, usando el LRT, hemos identificado que las variables clínico-patológicas sobreexpresión del HER2, subtipos IHQ y subtipo molecular determinado por PAM50, están significativamente relacionados con la pCR. Según el modelo univariante, la sobreexpresión del HER2 aumenta en casi 4 veces la probabilidad de conseguir pCR. Esos subtipos determinados por IHQ como HER2+ o TN aumenta en 5 y 2 veces, respectivamente, la probabilidad de conseguir pCR cuando comparamos con los subtipos Luminales, y en cuanto a los subtipos moleculares definidos por PAM50, el HER2enriched y Basal-like (No-luminales) tienen más de 2.5 veces la probabilidad de conseguir pCR que los Luminales (Luminal A y Luminal B). En cuanto a los genes DE, todos están relacionados de manera significativa con la pCR exceptuando el gen *GNG4*. Estos resultados son coherentes con el análisis de expresión diferencial entre las pacientes con pCR y sin ella. El OR de los genes DE relacionados con pCR obtenidos en el modelo univariante también siguen la tendencia de los log₂ FC obtenidos en el AED. Aquellos genes con el OR mayor de 0, relacionados de manera positiva con pCR, son aquellos que se sobreexpresan en las pacientes con pCR en el AED, como es el caso de *CDC6*, *FGFR4*, *MYCN*, *IDH2* y *RASGRF1*. Y aquellos el OR menores de 0, están expresados a la baja en las pacientes con pCR, como es el caso de *LEFTY1*, *STK11*, *FGF10* y *WNT3* (Tabla 11).

Tabla 11. Análisis de regresión logística univariante entre las variables clínico-patológicas incluidas en este estudio y los 10 genes DE y la variable pCR, Odds Ratio y p-valor del *Likelihood Ratio test* (LRT) de las relaciones. Solo se muestran las relaciones con un p-valor < 0.05.

Variables relacionadas con pCR	Odds Ratio (95% CI)		p-valor del LRT
Sobreexpresión de HER2			0.008
Positivo vs negativo	3.9	(1.4 - 11.34)	
Subtipos IHQ			0.014
HER2+ vs Luminal	5.33	(1.7 - 17.83)	
Triple negativo vs Luminal	2.31	(0.66 - 6.91)	
Subtipos moleculares por PAM50			0.040
No-Luminal vs Luminal	2.68	(1.04 - 7.42)	
CDC6	1.65	(1.158 - 2.52)	0.004
LEFTY1	0.753	(0.60 - 0.91)	0.0047
FGFR4	1.48	(1.20 - 1.92)	4.03E-05
MYCN	1.22	(1.02 - 1.47)	0.02
IDH2	2.22	(1.25 - 4.4)	0.0053
STK11	0.11	(0.026 - 0.39)	0.00021
FGF10	0.79	(0.63 - 0.98)	0.035
RASGRF1	1.24	(1.06 - 1.48)	0.0072
WNT3	0.75	(0.56 - 0.97)	0.03

Para asegurarnos de que la información de los genes DE aportaban valor predictivo (y no solo igualan) a las variables ya conocidas, o incluso aportan más información por ellos mismos, se dividen en 3 sets de datos, por un lado, el **set CLIN** que incluye solo las variables clínico-patológicas inclusive los biomarcadores de pCR ya conocidos, el **set GENEX** que incluye solo la información de los genes DE, y el **set CLIN_GENEX** que incluye tanto la información clínica como la información genética.

Con el **análisis predictivo multivariante** basado en LASSO se han generado tres modelos predictivos para cada uno de los sets de datos (CLIN, GENEX y CLIN_GENEX) en las diferentes *Cohortes_Discovery* (Cohorte_D_Allsubtypes, Cohorte_D_HER2- y Cohorte_D_ER+). Aparte de estos 3 modelos generados en base a los 3 sets de datos, hemos incluido un modelo que solo incluye a la variable **PAM50** para ver el valor predictivo que tiene en nuestra *Cohorte_Discovery*.

Según su peso en el modelo, se han seleccionado las variables más importantes para la predicción de pCR en cada uno de los sets de datos (CLIN, GENEX y CLIN_GENEX). Sólo se han seleccionado aquellas variables con menos del 5% con coeficiente igual a 0 y esas serán las que compongan el modelo. Una vez definidos los modelos, la evaluación de ellos se realizará determinando el área bajo la curva ROC y AUC mediante la representación gráfica de la sensibilidad de nuestro modelo frente a la especificidad de este. El valor de dicho AUC se puede interpretar como la probabilidad de que un clasificador puntúe una instancia positiva aleatoriamente más alta que la negativa. Cuanto más lejana esté la línea de un modelo predictivo de la línea de estimación aleatoria de la curva ROC mayor poder predictivo tiene. Como el modelo PAM50 incluye sólo a la variable PAM50 no hay necesidad de seleccionar variables en base al % con coeficiente igual a 0. Los resultados se representan mediante: un *barchart*, que nos identifica el número de veces que la variable es seleccionada por Lasso durante las repeticiones de la validación externa por LOO, *crossbars*, que muestra el coeficiente de cada variable de nuestro modelo, los *violin plots*, que nos permiten observar gráficamente la distribución de los AUCs obtenidos durante las 1000 iteraciones, y los *violin ratios*, que nos da una idea de la mejoría de un modelo con respecto a otro.

4.4.1.3.1. Predicción de pCR en la Cohorte_D_Allsubtypes

En el modelo *CLIN_Allsubtypes* la variable seleccionada ha sido la sobreexpresión del HER2; en el modelo *GENEX_Allsubtypes*, la expresión de los genes *CDC6*, *FGFR4*, *LEFTY1*, *STK11* e *IDH2*. La combinación de la sobreexpresión del HER2 y de la afectación ganglionar clínica, más la información de la expresión génica de los genes *FGFR4*, *LEFTY1*, *STK11*, *CDC6* e *IDH2*, forman el modelo predictivo *CLINGENEX_Allsubtypes* (Figura 18).

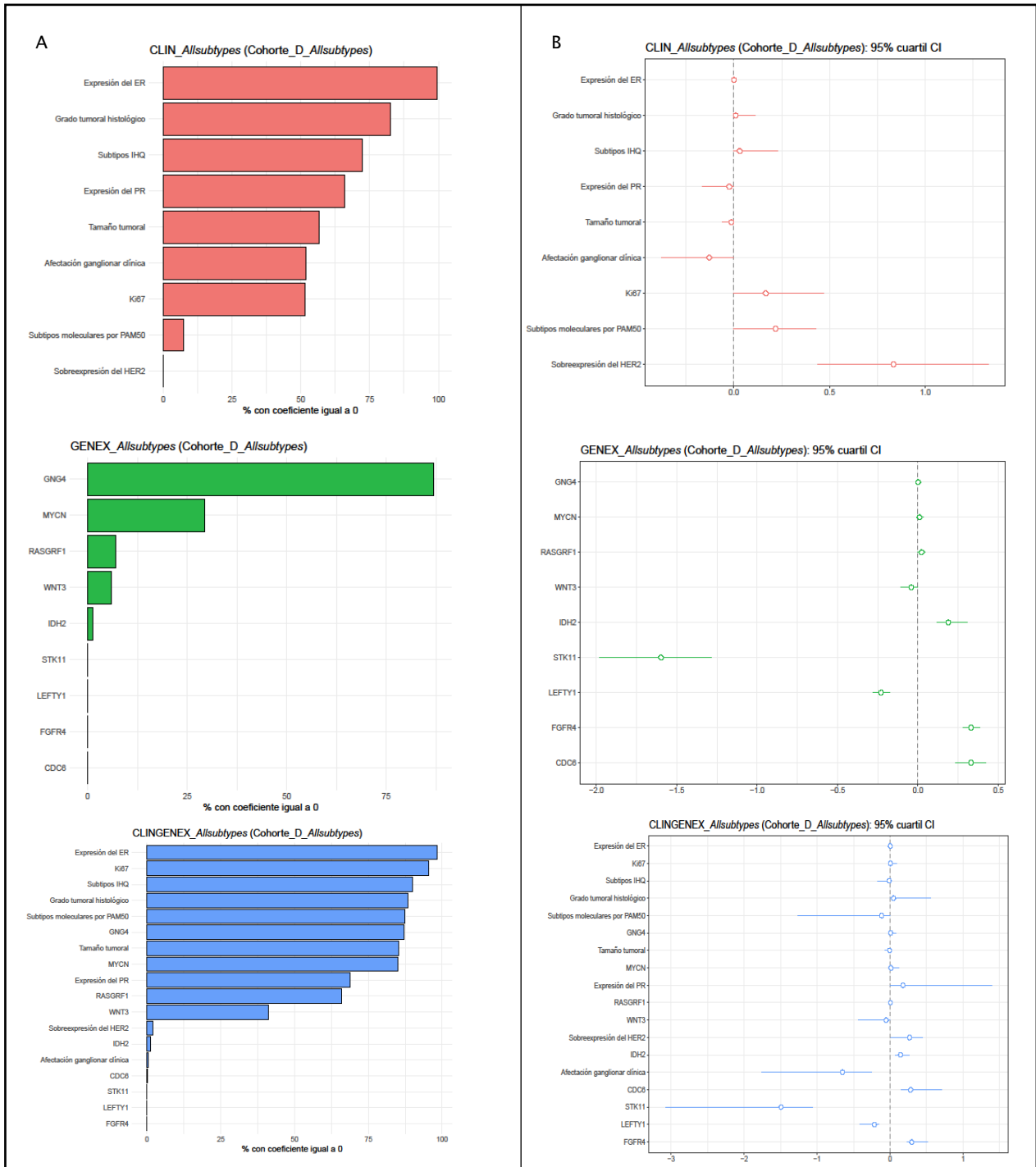


Figura 18. Representación de la importancia de las variables dentro de cada uno de los modelos predictivo de pCR de la Cohorte_D_AllSubtypes. A. Diagrama de barras donde se muestran las variables con más peso dentro de los tres sets de datos (CLIN, GENEX, CLINGENEX). En el eje de ordenadas se identifican las variables en orden según la importancia dentro del modelo, en el eje de abscisa se muestra el porcentaje del número de veces que la variable seleccionada por LASSO obtiene un coeficiente 0. B. Crossbars. Se identifica el coeficiente de cada variable. Al haberse iterado mil veces, se identifica el intervalo de confianza del 95%. El punto representa la media y la línea representa el intervalo.

En este caso, y tal y como se muestra en la representación gráfica de los modelos de este apartado (Figura 19), tanto el modelo PAM50 como el modelo *CLIN_Allsubtypes* obtienen los valores más bajos de AUC (0.51 y 0.55, respectivamente) y sus curvas tienen menos distancia a la estimación aleatoria, por lo que su valor predictivo no es muy diferente al propio azar. Cuando la información de la expresión génica es combinada con la información clínica, como es el caso del modelo *CLINGENEX_Allsubtypes*, el valor de nuestro modelo se aleja de la estimación aleatoria y el valor del AUC aumenta hasta el 0.74. En este caso, la información de la expresión génica por sí sola (modelo *GENEX_Allsubtypes*) es capaz de predecir la respuesta al tratamiento con el valor de AUC más alto, 0.79. En el *violin plot*, la forma de violín identifica la distribución que siguen dichos AUCs; de esta forma, en el modelo *GENEX_Allsubtypes* vemos que el valor se encuentra muy centrado y compacto en el 0.79, mientras que en el modelo *CLIN_Allsubtypes* o PAM50 tienen una distribución más extendida o lo que es lo mismo, el rango del AUC es mayor. En la Tabla 12 se resumen los modelos de este apartado con las variables seleccionadas y el AUC estimado.

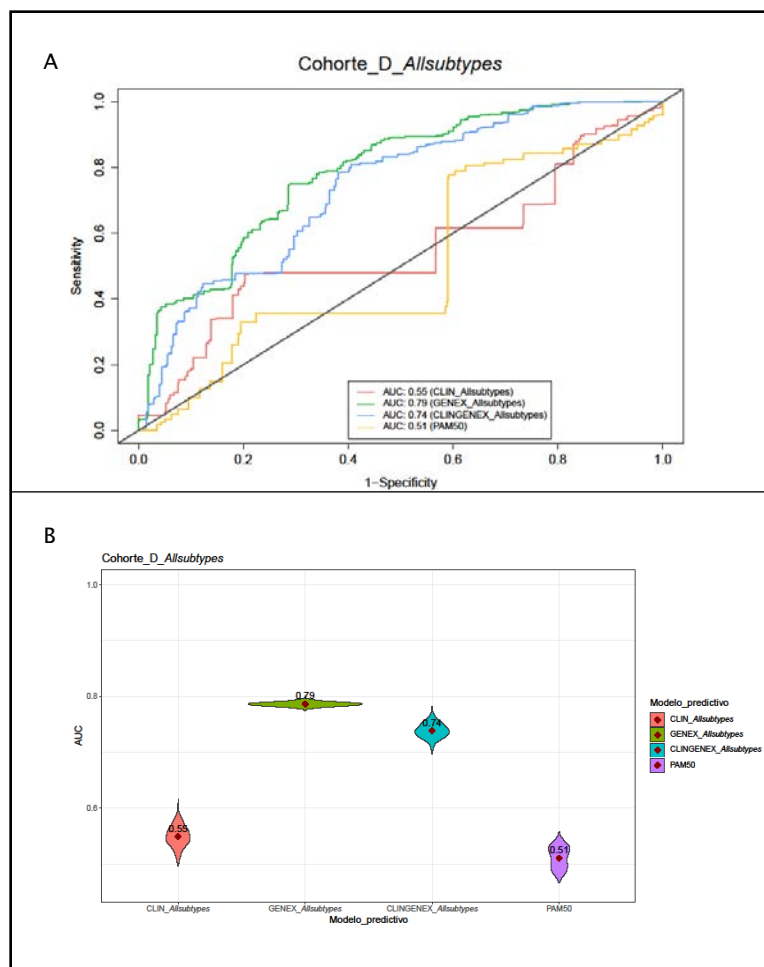


Figura 19. A. Curvas ROC de los modelos predictivos de pCR de la Cohorte_D_Allsubtypes. B. Los *violin plots* representan la distribución del AUC obtenido durante las mil iteraciones.

Tabla 12. Resumen de los modelos predictivos de la cohorte_D_Allsubtypes con las variables seleccionadas (aquellas con menos del 5% de coeficiente igual a 0 en la selección con LASSO) y el AUC obtenido en cada caso.

Cohorte	Modelo predictivo	Variables seleccionadas	AUC
Cohorte_D_Allsubtypes	CLIN_Allsubtypes	Sobreexpresión de HER2	0.55
	GENEX_Allsubtypes	CDC6 + FGFR4 + LEFTY1 + STK11 + IDH2	0.79
	CLINGENEX_Allsubtypes	Sobreexpresión de HER2 + Afectación ganglionar clínica + FGFR4 + LEFTY1 + STK11 + IDH2	0.74
	PAM50	PAM50	0.51

4.4.1.3.2. Predicción de pCR en la Cohorte_D_HER2-

En el modelo CLIN_HER2- las variables seleccionadas han sido el grado tumoral histológico y el tamaño tumoral mamográfico, en el modelo GENEX_HER2-, la expresión de los genes *STK11*, *LEFTY1* y *FGFR4*. La combinación del tamaño tumoral mamográfico, la afectación ganglionar clínica y la expresión del ER por IHQ más la información de la expresión génica de los genes *FGFR4*, *LEFTY1* y *STK11* forman el modelo predictivo CLINGENEX_HER2- (Figura 20).

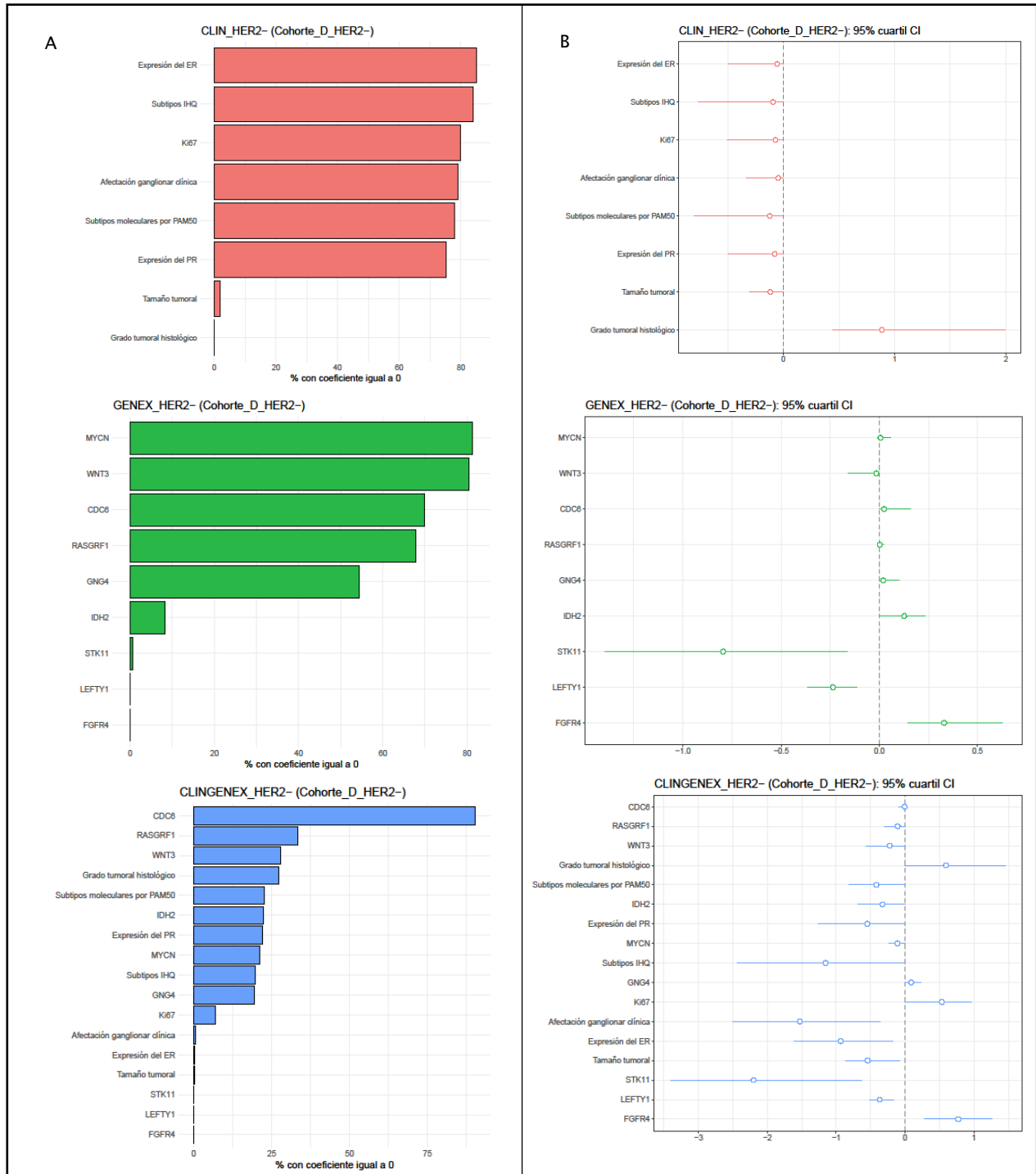


Figura 20. Representación de la importancia de las variables dentro de cada uno de los modelos predictivo de pCR de la Cohorte_D_HER2-. A. Diagrama de barras donde se muestran las variables con más peso dentro de los tres sets de datos (CLIN, GENEX, CLINGENEX). En el eje de ordenadas se identifican las variables en orden según la importancia dentro del modelo, en el eje de abscisa se muestra el porcentaje del número de veces que la variable seleccionada por LASSO obtiene un coeficiente 0. B. *Crossbars*. Se identifica el coeficiente de cada variable. Al haberse iterado mil veces, se identifica el intervalo de confianza del 95%. El punto representa la media y la línea representa el intervalo.

Tal y como se muestra en la representación gráfica de los modelos de este apartado (Figura 21), tanto el modelo PAM50 como el modelo CLIN_HER2- obtienen los valores más bajos de AUC (0.52 y 0.6, respectivamente) y sus curvas tienen menos distancia a la estimación aleatoria, por lo que su valor predictivo no es muy diferente al propio azar. Cuando la información de la expresión génica es combinada con la información clínica, como es el caso del modelo CLINGENEX_HER2-, el valor predictivo de nuestro modelo se aleja de la estimación aleatoria y el valor del AUC aumenta hasta el 0.85, siendo el modelo con el mayor AUC de aquellas pacientes. La información de la expresión génica por sí sola (modelo GENEX_HER2-) es capaz de predecir la respuesta al tratamiento con un valor de AUC de 0.83, que aun no siendo el AUC más alto, está muy cercano a él además de tener un comportamiento muy similar en la curva ROC. En el *violin plot* la forma de violín identifica la distribución que siguen dichos AUCs. En el modelo GENEX_HER2-, y algo menos en el CLINGENEX_HER2-, vemos que los valores se encuentran muy centrados y compactos en el 0.83 y 0.85, mientras que en el modelo CLIN_HER2- o PAM50 tienen una distribución más extendida. En la Tabla 13 se resumen la información de los modelos predictivos de este apartado.

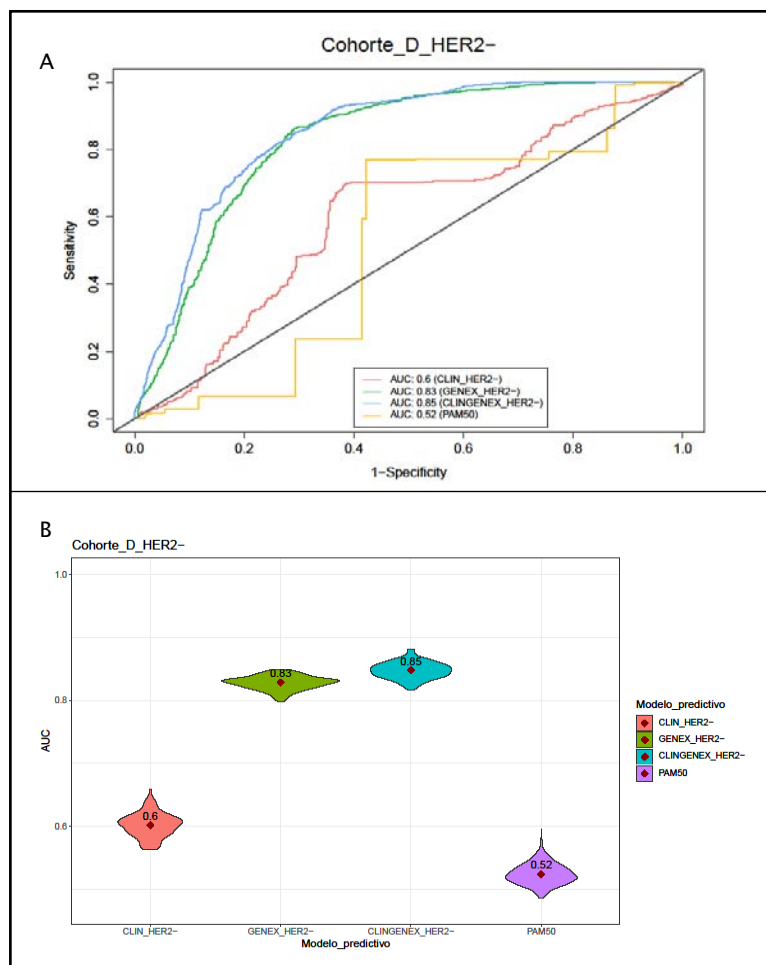


Figura 21. A. Curvas ROC de los modelos predictivos de pCR de la Cohorte_D_HER2-. B. Los *violin plots* representan la distribución del AUC obtenido durante las mil iteraciones.

Tabla 13. Resumen de los modelos predictivos de la cohorte_D_HER2- con las variables seleccionadas (aquellas con menos del 5% de coeficiente igual a 0 en la selección con LASSO) y el AUC obtenido en cada caso.

Cohorte	Modelo predictivo	Variables seleccionadas	AUC
Cohorte_D_HER2-	CLIN_HER2-	Grado tumoral histológico + tamaño tumoral mamográfico	0.6
	GENEX_HER2-	STK11 + LEFTY1 + FGFR4	0.83
	CLINGENEX_HER2-	Tamaño tumoral mamográfico + afectación ganglionar clínica + ER + FGFR4 + LEFTY1 + STK11	0.85
	PAM50	PAM50	0.52

4.4.1.3.3. Predicción de pCR en la Cohorte_D_ER+

En el modelo CLIN_ER+ las variables seleccionadas han sido el grado tumoral histológico, el tamaño tumoral mamográfico y la afectación ganglionar clínica, en el modelo GENEX_ER+, la expresión de los genes *CDC6*, *FGFR4*, *RASGRF1*, *LEFTY1*, *WNT3* y *GNG4*. La combinación del grado tumoral histológico, el tamaño tumoral mamográfico, la afectación ganglionar y la expresión de PR determinada por IHQ más la información de la expresión génica de los genes *CDC6*, *FGFR4*, *LEFTY1*, *WNT3* y *GNG4* forman el modelo predictivo CLINGENEX_ER+ (Figura 22).

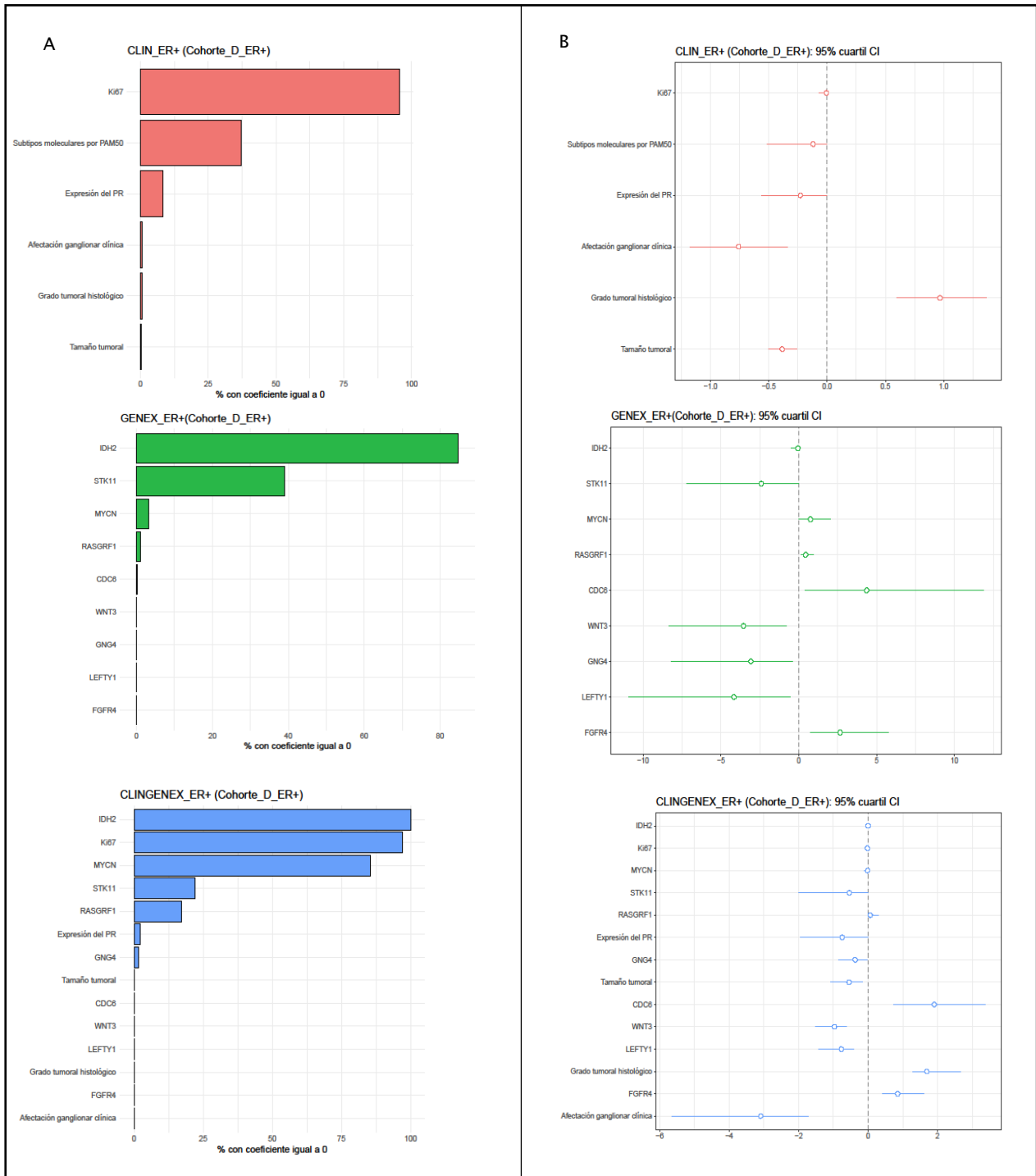


Figura 22. Representación de la importancia de las variables dentro de cada uno de los modelos predictivo de pCR de la Cohorte_D_ER+. A. Diagrama de barras donde se muestran las variables con más peso dentro de los tres sets de datos (CLIN, GENEX, CLINGENEX). En el eje de ordenadas se identifican las variables en orden según la importancia dentro del modelo, en el eje de abscisa se muestra el porcentaje del número de veces que la variable seleccionada por LASSO obtiene un coeficiente 0. B. *Crossbars*. Se identifica el coeficiente de cada variable. Al haberse iterado mil veces, se identifica el intervalo de confianza del 95%. El punto representa la media y la línea representa el intervalo.

El modelo CLIN_ER+ obtiene el valor más bajo de AUC, con un 0.71. En los modelos CLINGENEX_ER+ y GENEX_ER+ los valores de AUC están muy cercanos a la perfección de un clasificador con valores cercanos al 1, ambos modelos obtienen un valor AUC de

0.96. El valor de nuestro modelo se aleja de la estimación aleatoria y el valor del AUC aumenta hasta el 0.74. En este caso, la información de la expresión génica por si sola (modelo GENEX_ER+) es capaz de predecir la respuesta al tratamiento con el valor de AUC más alto, 0.79. En el violin plot la forma de violín identifica la distribución que siguen dichos AUCs, de esta forma, en el modelo GENEX_ER+ vemos que el valor se encuentra muy centrado y compacto en el 0.79, mientras que en el modelo CLIN_ER+ o PAM50 tienen una distribución más extendida o lo que es lo mismo, el rango del AUC es mayor (Figura 23). En la Tabla 14 se muestran las variables seleccionadas para cada uno de los modelos y el resultado de AUC.

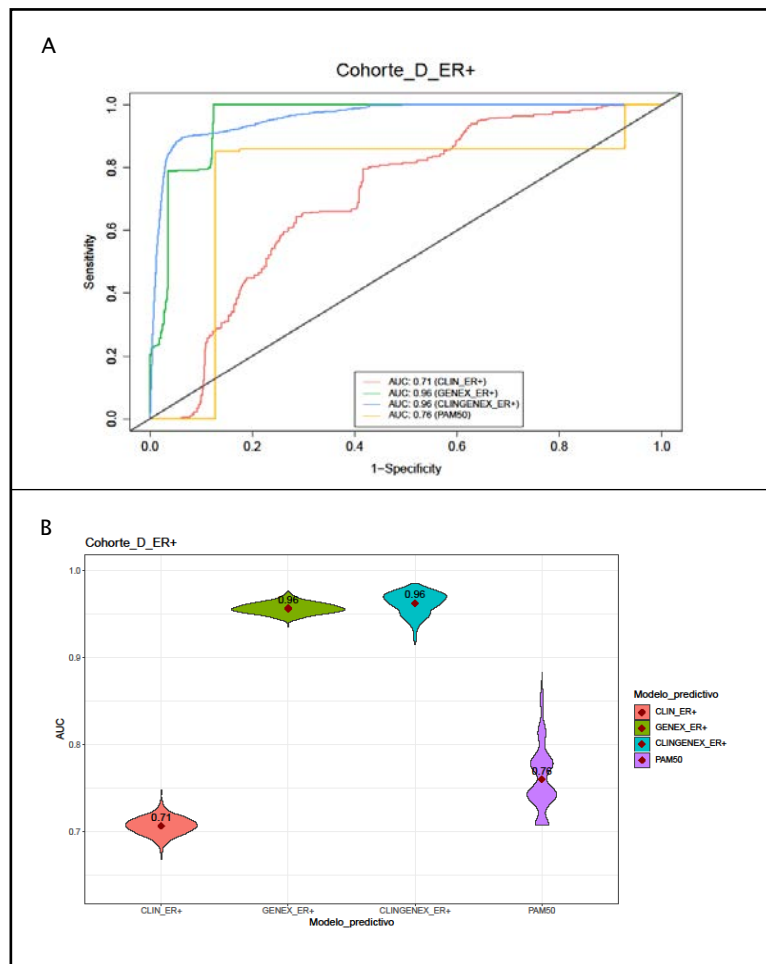


Figura 23. A. Curvas ROC de los modelos predictivos de pCR de la Cohorte_D_ER+. B. Los *violin plots* representan la distribución del AUC obtenido durante las mil iteraciones.

Tabla 14. Resumen de los modelos predictivos de la cohorte_D_ER+ con las variables seleccionadas (aquellas con menos del 5% de coeficiente igual a 0 en la selección con LASSO) y el AUC obtenido en cada caso.

Cohorte	Modelo predictivo	Variabes seleccionadas	AUC
Cohorte_D_ER+	CLIN_ER+	Grado tumoral histológico + tamaño tumoral mamográfico + Afectación ganglionar clínica	0.71
	GENEX_ER+	CDC6 + FGFR4 + RASGRF1 + LEFTY1 + WNT3 + GNG4	0.96
	CLINGENEX_ER+	Grado tumoral histológico + Tamaño tumoral mamográfico + Afectación ganglionar clínica + PR + CDC6 + FGFR4 + LEFTY1 + WNT3 + GNG4	0.96
	PAM50	PAM50	0.76

4.4.1.3.4. Mejora de los modelos predictivos frente a PAM50 como predictor individual

En los *violíns plots* (Figura 24) donde se muestran las ratios de los modelos predictivos de pCR que proponemos frente al modelo PAM50, observamos que sólo en aquellos casos donde **no** se incluye la información molecular, PAM50 tiene mayor o prácticamente igual poder predictivo. Cuando la expresión de los genes específicos de cada modelo de cada cohorte es incluida, PAM50 pierde valor en la determinación de la respuesta al tratamiento con NAC, siendo más pronunciado en pacientes que incluyen todos los tipos de tumores de mama y en los tumores HER2-. No es tan acusada, aunque sigue siendo superior, en los tumores HER2-/ER+.

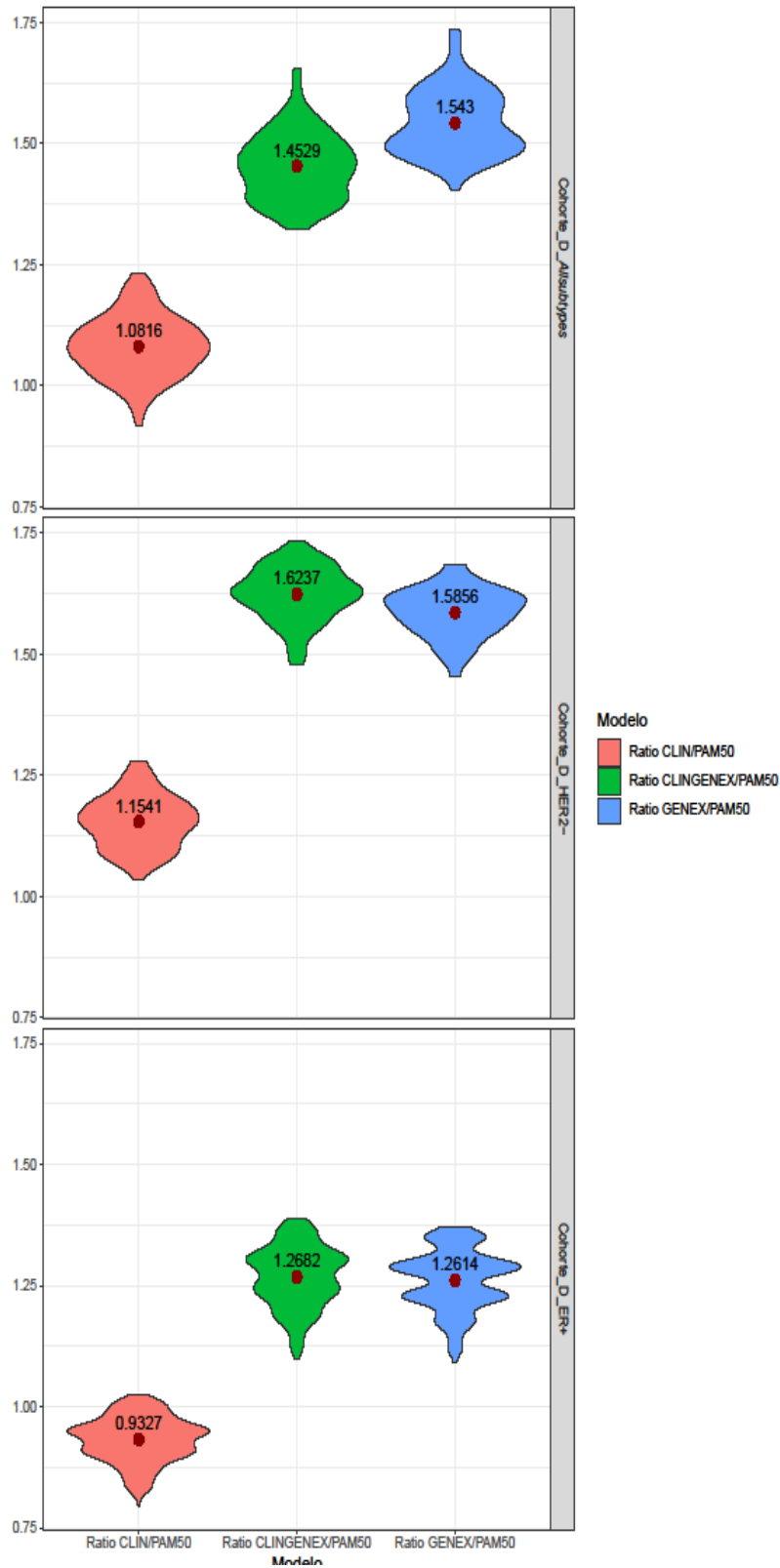


Figura 24. Violin plots. Se representa la ratio del AUC de los modelos definidos en las diferentes cohortes versus el modelo que incluye solo la variable PAM50. Valores por encima de 1 indican una mejoría de los modelos propuestos frente al modelo PAM50.

4.4.1.3.5. Validación de los modelos predictivos

Tal y como se ha especificado en la sección material y métodos, los modelos predictivos que han sido definidos en la Cohorte_ *Discovery* han sido evaluados en cohortes externas que tienen el mismo esquema de tratamiento y con la información clínico-patológica y molecular necesaria que requieren los modelos. Este no es el caso de la variable PAM50, que en las cohortes externas que han sido exploradas para la validación de nuestros modelos no está presente, por lo que **el predictor PAM50 no queda contemplado en la validación** de nuestros resultados. En cuanto a la división de las cohortes según su estatus de HER2 y ER, la cohorte de validación (Cohorte_ *Validation*) ha sido dividida de la misma forma que la cohorte de descubrimiento, con tres cohortes **Cohorte_V_Allsubtypes/HER2-/ER+**. En las curvas ROC podemos observar cómo se comporta cada uno de los modelos con respecto a la línea de aleatoriedad. En el caso de la cohorte que incluye todos los subtipos de cáncer de mama (Cohorte_V_ *Allsubtypes*), vemos que el modelo con mayor poder predictivo, y por tanto más alejado de la línea, es el modelo que incluye tanto información clínico-patológica como molecular, el modelo CLINGENEX_ *Allsubtypes* con un AUC de 0.74. En el violins plot, observamos que el rango que AUCs es menor en el modelo CLIN_ *Allsubtypes* encontrando en su límite inferior valores por debajo del 0.5. Si comparamos el rango de los modelos con los AUCs más alto, GENEX_ *Allsubtypes* vs CLINGENEX_ *Allsubtypes*, el modelo más compacto en torno a su AUC es el modelo CLINGENEX_ *Allsubtypes*. En la cohorte que sólo incluye a aquellas pacientes con tumores HER2- (independientemente del estatus de ER), no existen grandes diferencias entre los tres modelos predictivos y en este caso la información molecular por si sola (GENEX_HER2-) no parece ser más informativa que las variables clínico-patológicas (CLIN_HER2-) ya que el AUC es menor (0.58 vs 0.63, respectivamente). Sin embargo, la combinación de ambas (CLINGENEX_HER2-) se posiciona como el mejor modelo para predecir la respuesta al tratamiento con NAC en pacientes HER2- (AUC 0.67). De la Cohorte_V_ER+ que incluye aquellos tumores con HER2- y ER+, los modelos CLIN_ER+ y GENEX_ER+ por separado obtienen el mismo poder predictivo o valor de AUC (0.71) mientras que el modelo CLINGENEX_ER+ tiene el mayor AUC o poder predictivo (0.76), Figura 25.

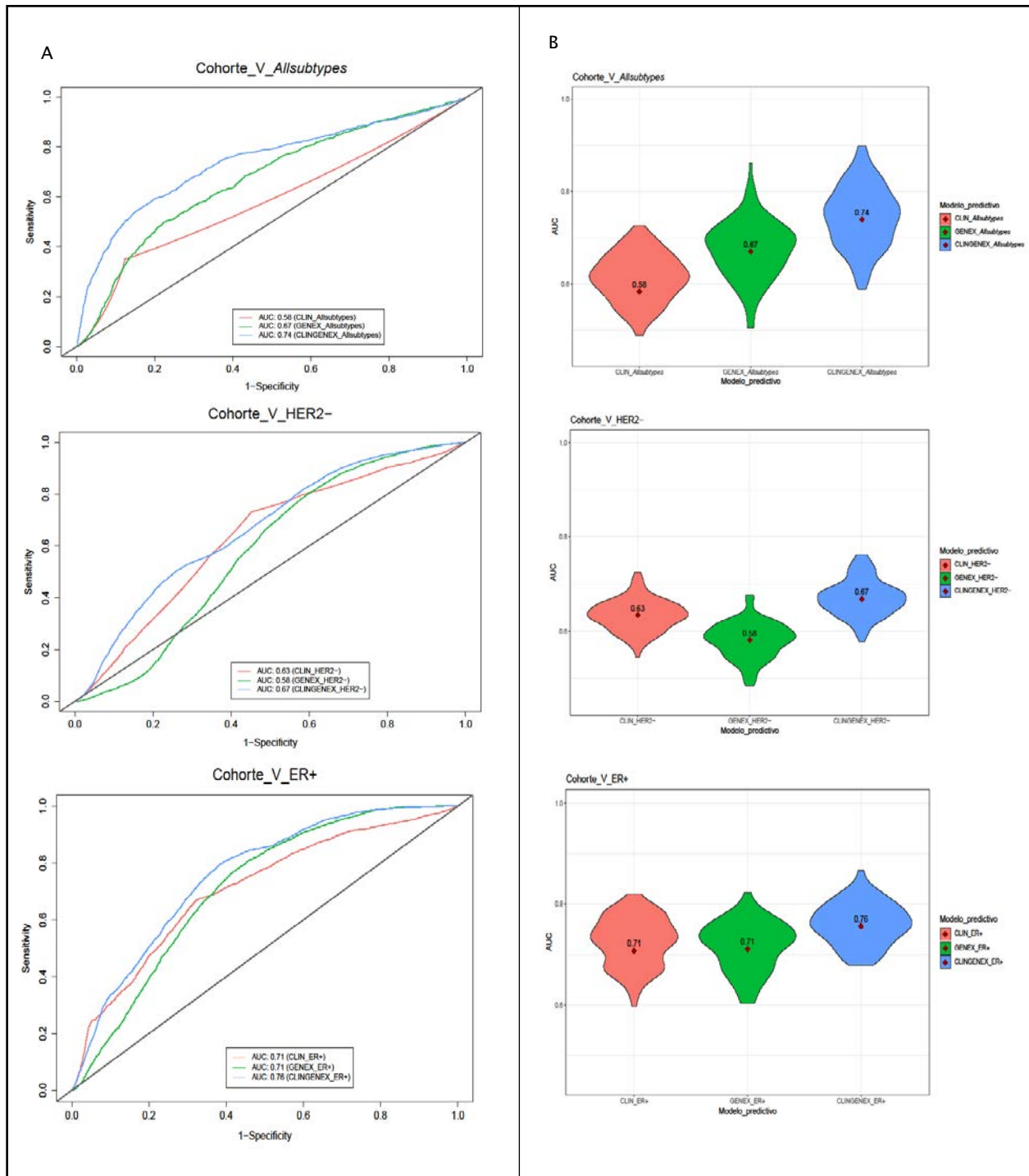


Figura 25. Validación de los modelos predictivos CLIN_*Allsubtypes*/HER2-/ER+, GENEX_*Allsubtypes*/HER2-/ER+ y CLINGENEX_*Allsubtypes*/HER2-/ER+ en las Cohortes_Validation. A. Curvas ROC de los modelos predictivos de pCR en las Cohortes_Validation. B. Los violin plots representan la distribución del AUC obtenido durante las mil iteraciones de cada uno de los modelos en las diferentes Cohortes_Validation.

En la Tabla 15 han sido resumidos todos los AUCs obtenidos de todos los modelos en ambas cohortes, tanto la de descubrimiento como la de validación. De las cohortes de pacientes con todos los subtipos de cáncer de mama, el mejor modelo es el GENEX_*Allsubtypes*. En la cohorte de validación pierde fuerza predictiva, aunque sigue mejorando al modelo CLIN_*Allsubtypes*. Para los tumores HER2-, el mejor modelo para predecir la respuesta al tratamiento es el que combina tanto la información clínica como la información molecular (CLINGENEX_HER2-), afirmación que se valida en la cohorte externa. Y lo mismo ocurre con la cohorte de pacientes HER2-/ER+, para los que, aunque el modelo GENEX_ER+ y CLINGENEX_ER+ muestran el mismo valor predictivo (AUC 0.96), los resultados de la validación otorgan al modelo CLINGENEX_ER+ el máximo valor predictivo.

Tabla 15. Resumen de todos los AUCs obtenidos de los modelos predictivos tanto en las cohortes donde han sido diseñadas como en la cohorte de validación (incluyendo PAM50 en las Cohortes_*Discovery*).

	Cohortes	Modelo predictivo	AUC
Cohorte_ <i>Discovery</i>	Cohorte_D_ <i>Allsubtypes</i>	CLIN_ <i>Allsubtypes</i>	0.55
		GENEX_ <i>Allsubtypes</i>	0.79
		CLINGENEX_ <i>Allsubtypes</i>	0.74
		PAM50	0.51
	Cohorte_D_HER2-	CLIN_HER2-	0.6
		GENEX_HER2-	0.83
		CLINGENEX_HER2-	0.85
		PAM50	0.52
	Cohorte_D_ER+	CLIN_ER+	0.71
		GENEX_ER+	0.96
		CLINGENEX_ER+	0.96
		PAM50	0.76
Cohorte_ <i>Validation</i>	Cohorte_V_ <i>Allsubtypes</i>	CLIN_ <i>Allsubtypes</i>	0.58
		GENEX_ <i>Allsubtypes</i>	0.67
		CLINGENEX_ <i>Allsubtypes</i>	0.74
	Cohorte_V_HER2-	CLIN_HER2-	0.63
		GENEX_HER2-	0.58
		CLINGENEX_HER2-	0.67
	Cohorte_V_ER+	CLIN_ER+	0.71
		GENEX_ER+	0.71
		CLINGENEX_ER+	0.76

4.1.4.2 Análisis de supervivencia

Hemos definido la supervivencia en base al evento MCE. De los 84 pacientes que hemos considerado en esta parte del estudio, 44 personas han sobrevivido (independientemente si actualmente tienen enfermedad o no), 27 han fallecido con la enfermedad y de 13 no disponemos de la información. En la Figura 26 observamos la curva de Kaplan y Meier con la supervivencia global, enfrentando la probabilidad de supervivencia frente al tiempo en días. En nuestra población de estudio hay muchos pacientes que no sufren el evento, es decir, muerte con enfermedad, siendo la probabilidad de que un paciente sobreviva al menos un 50% de 3413 días.

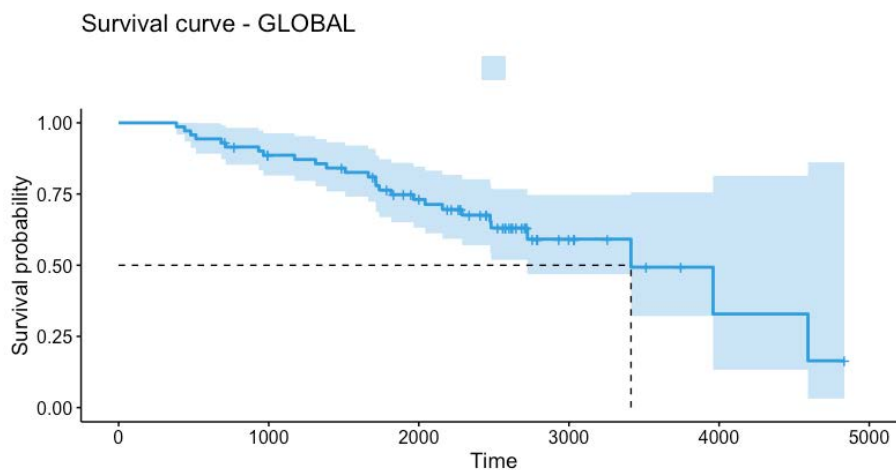


Figura 26. Curva de Supervivencia de Kaplan–Meier de las pacientes incluidas en este estudio teniendo como evento la muerte con enfermedad (MCE), en este caso cáncer de mama o metástasis derivado de él.

En la Figura 27 se muestran las curvas de Kaplan y Meier de las variables categóricas significativas. La aplicación de un *Log Rank test*, nos permite distinguir si las curvas son distintas para algunos de los grupos. En nuestro estudio variables como la expresión del ER, el subtipo determinado por PAM50, el porcentaje de Ki67 y la expresión génica de *FGF10* y de *IDH2* están relacionadas con la supervivencia. En el caso del marcador ER, la expresión positiva de este proporciona al paciente una mayor probabilidad de supervivencia ($p=0.022$). El Basal–Like y HER2–enriched son los subtipos de PAM50 que menor supervivencia presentan, mientras que los Luminales, sobre todo el Luminal B, están relacionados con una mayor supervivencia ($p=0.029$), aquellos pacientes con tumores más proliferativos, según el marcador Ki67 antes del tratamiento con NAC,

tienen menos probabilidad de supervivencia que aquellos tumores menos proliferativos ($p=0.0022$). Agrupando la expresión de los genes en 3 cuartiles, alta expresión, expresión media y baja expresión, se observa que tener una alta o media expresión del gen *FGF10* proporciona una mayor supervivencia que tener una expresión baja y además es la variable con mayor poder estadístico en la diferencia entre sus curvas de supervivencia ($p<0.0001$). El contrario es el caso de la expresión génica de *IDH2*, donde una alta expresión se relaciona con probabilidad de supervivencia menor ($p=0.037$).

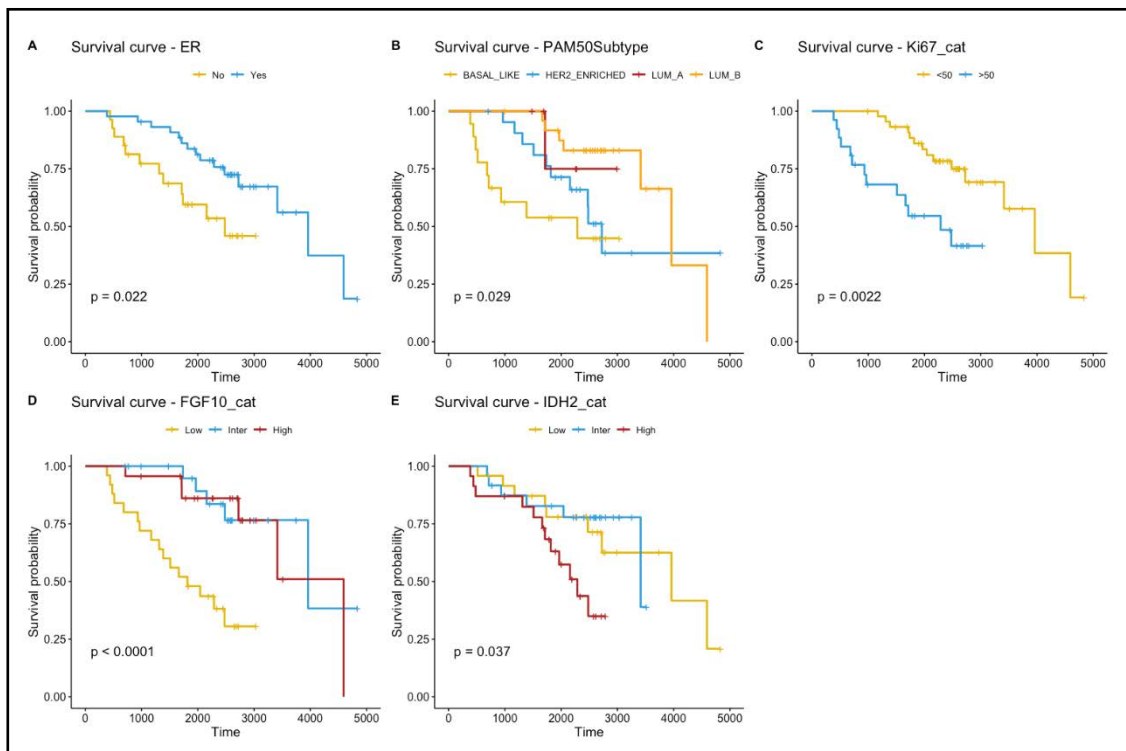
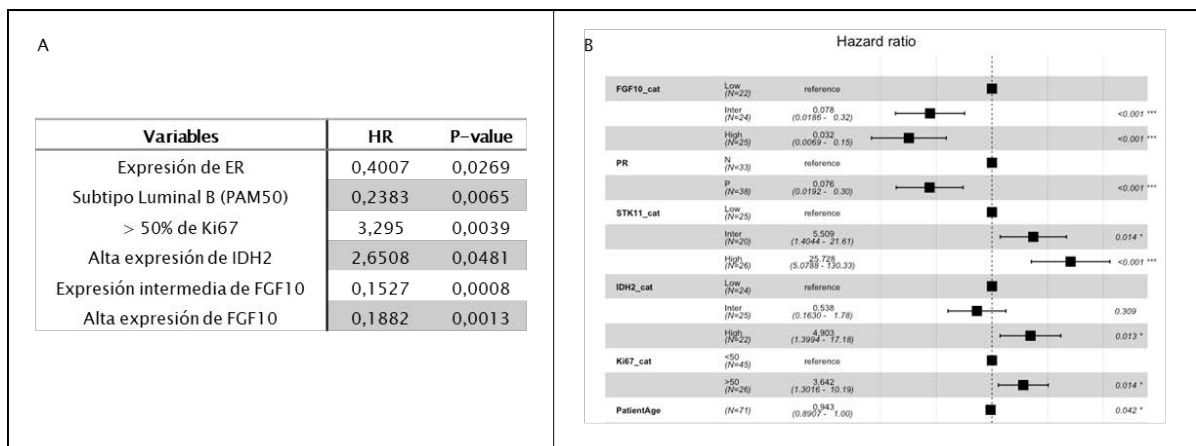


Figura 27. Se muestran las curvas de supervivencia Kaplan-Meier de las variables que tienen una relación estadísticamente significativa con el evento MCE. La comparación entre categorías dentro de la misma variable se ha realizado con una prueba no paramétrica (Log Rank). A. Comparación entre la categoría ER- y ER+ de la variable ER. B. Curvas de las 4 categorías dentro del subtipos de PAM50. C. Diferencia entre las curvas de supervivencia entre un porcentaje de Ki67 mayor y menor de 50. D y E. La expresión de los genes *FGF10* y *IDH2* se ha dividido en 3 cuartiles para determinar las diferencias entre sus curvas de supervivencia.

Los modelos de regresión univariante y multivariante de COX (Tabla 16) para el estudio de supervivencia se han llevado a cabo con la información clínica y de expresión génica. En el análisis univariante, las variables con relación estadísticamente significativa con la supervivencia son: la expresión positiva de ER, el subtipo Luminal B determinado por PAM50, el Ki67 mayor del 50%, una alta expresión de *IDH2*, y una alta e intermedia expresión de *FGF10*. Según el HR de cada una de las variantes de las variables, la

expresión positiva de ER, el subtipo Luminal B y niveles de expresión altos e intermedios de *FGF10* parecen conferir protección frente al evento, en este caso muerte con MCE. Sin embargo, valores de Ki67 por encima del 50% y alta expresión del gen *IDH2* aumentan la probabilidad de dicho evento. En el análisis multivariante se mantiene la disminución de la probabilidad de MCE en aquellos pacientes que tienen alta e intermedia expresión en *FGF10*. La expresión de PR y la edad del paciente también se incluyen como factores protectores del evento. También, como en el caso del univariante, la alta expresión de *IDH2*, y porcentajes mayores al 50 en Ki67 están relacionados el aumento de MCE. Muy significativo es el caso de la alta expresión de *STK11* siendo la probabilidad de que ocurra el evento 25,7 veces superior que en el caso de una expresión a la baja de este gen ($p < 0.001$).

Tabla 16. Resultados de los modelos de regresión logística de COX con todas las variables incluidas en este estudio, solo se muestran aquellas variables donde en alguna de sus categorías tiene una relación estadísticamente significativa con el evento MCE. A. Resultados del análisis univariante de COX. B. Resultados del análisis multivariante de COX.



4.4.2. Alteraciones genéticas (DNA-Cohorte)

4.4.2.1. Características clinicopatológicas

Las pacientes incluidas en esta parte del estudio son 25, con un tamaño tumoral de 3 cm de media, un 60% de las pacientes tenían afectación ganglionar en el diagnóstico y prevalece el grado tumoral histológico entre 1 y 2 (más del 50% de las pacientes). Más del 50% de las pacientes expresan el ER y un 24% sobreexpresan el HER2. En esta cohorte, la mayoría de las pacientes tienen un porcentaje de Ki67 menor de 50 y prevalece los subtipos Luminales tanto definidos por IHQ como por PAM50. La distribución de las pacientes que han obtenido pCR frente a las que no está muy

desbalanceada (16 vs 84%, respectivamente) (Tabla 17). Debido a esta desproporcionalidad los resultados y conclusiones producto del análisis de estas pacientes no serán tenidos en cuenta y solo se proporcionará una descriptiva de las mutaciones encontradas.

Tabla 17. Características clínico-patológicas de la DNA-cohorte.

Variables clinicopatológicas	n	%
Todos los casos	25	100
Tamaño tumoral (media)	3	-
Afectación ganglionar clínica		
Positiva	15	60
Negativa	9	36
NA	1	4
Grado tumoral histológico		
1-2	13	52
3	10	40
NA	2	8
Expresión del ER		
Positivo	13	52
Negativo	12	48
Expresión del PR		
Positivo	12	48
Negativo	13	52
Sobreexpresión del HER2		
Positivo	6	24
Negativo	19	76
Ki67		
< 50%	14	56
> 50%	10	40
NA	1	4
Subtipos IHQ		
Luminales (A+B)	12	48
HER2+	6	24
Triples Negativos	7	28
Subtipos moleculares por PAM50		
Luminal A	1	4
Luminal B	9	36
HER2-enriched	6	24
Basal-Like	8	32
NA	1	4
Respuesta completa patológica (pCR)		
Si	4	16
No	21	84

4.4.2.2. Descriptiva de las mutaciones encontradas en T1

De las 25 pacientes de T1 secuenciadas, 23 de ellas tenían alguna alteración en los genes secuenciados. Han sido determinadas 82 mutaciones que se reparten entre los siguientes 31 genes: *TP53*, *ATM*, *RECQL4*, *BRCA2*, *ERBB2*, *AURKA*, *BRCA1*, *ERCC5*, *KMT2C*, *NF1*, *PIK3CA*, *POLE*, *TBX3*, *CCND3*, *CDH1*, *CDKN2A*, *EGFR*, *ESR1*, *FANCA*, *MAP3K1*, *MYC*, *PALB2*, *PTEN*, *ABCB1*, *ERCC2*, *FANCD2*, *KRAS*, *MAP2K4*, *MCL1*, *NCOR1* y *PIK3R1*. En la Tabla 18 se especifica el número de veces que un gen ha sido mutado (no se especifica si la mutación concreta) y en la Figura 28 se indica su porcentaje con respecto al total. El gen con más mutaciones es *TP53*, con el 16% de las mutaciones identificadas en él. Los genes *ATM*, *RECQL4* y *BRCA2* también se encuentran entre los genes más mutados (9, 6 y 5%, respectivamente).

Tabla 18. Recurrencia de los genes mutados en las pacientes incluidas en este estudio.

Genes Mutados en T1	Recurrencia del gen mutado
TP53	13
ATM	7
RECQL4	5
BRCA2	4
ERBB2	3
AURKA	3
BRCA1	3
ERCC5	3
KMT2C	3
NF1	3
PIK3CA	3
POLE	2
TBX3	3
CCND3	2
CDH1	2
CDKN2A	2
EGFR	2
ESR1	2
FANCA	2
MAP3K1	2
MYC	1
PALB2	2
PTEN	2
ABCB1	1
ERCC2	1
FANCD2	1
KRAS	1
MAP2K4	1
MCL1	1
NCOR1	1
PIK3R1	1
Total mutaciones determinadas	82

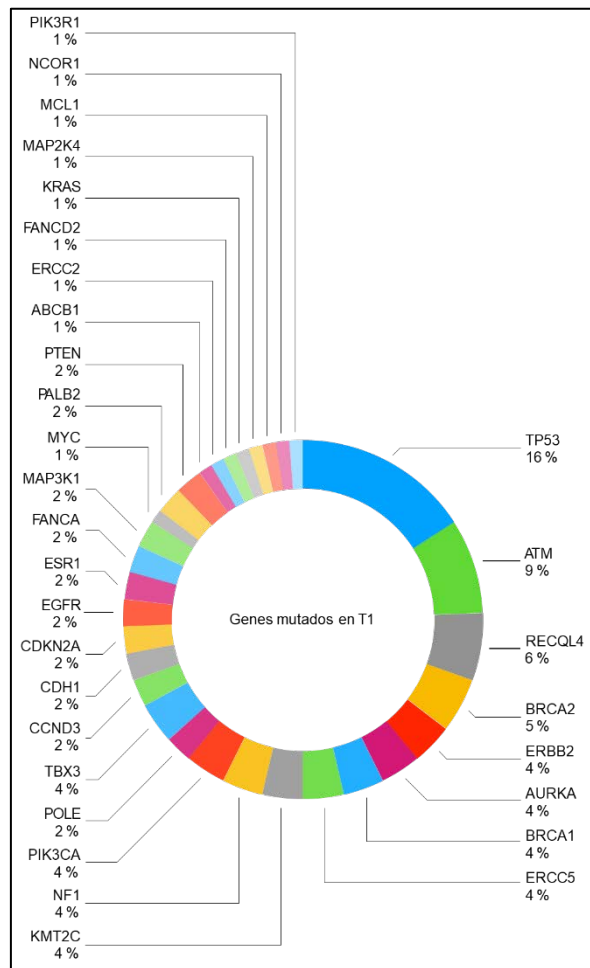


Figura 28. Porcentaje de los genes mutados.

En el *oncplot* A de la figura 29 se muestran los tipos de mutación encontrados en cada gen de las 23 pacientes con mutaciones. El tipo de mutación más común encontrado en los 31 genes es la mutación con cambio de sentido o *missense mutation*, un cambio puntual de la secuencia que da lugar a un codón diferente al original y que este equivale, por tanto, a un aminoácido diferente. El resultado son proteínas anómalas y, normalmente, no funcionales debido a los cambios. En todos los genes mutados, excepto *PIK3R1* y *MAP2K4*, encontramos mutaciones de cambio de sentido. El siguiente tipo de mutación más común es la mutación sin sentido o *nonsense mutation*, este tipo de mutación da lugar a un codón de parada que provoca la terminación prematura de la proteína. Encontramos *nonsense mutation* en los genes *ATM*, *BRCA2*, *NF1*, *MAP2K4* y *PIK3R1*. Desde el punto de vista del porcentaje de pacientes con mutaciones, el 52% de ellas tienen mutaciones en *TP53*, el 26% en *ATM*, el 17% en *BRCA2* y *RECQL4* y 13% de los genes *AURKA*, *BRCA1*, *ERCC5*, *KMT2C*, *NF1*, *PIK3CA* y *TBX3*. Las pacientes 45PNVV,

46PNVV y 49PNVV responden al tratamiento con NAC. La paciente 18MF tiene el mayor número de genes mutados, 13 y también el mayor número de mutaciones, en este caso 16. Esta paciente no presenta respuesta al tratamiento con NAC y el subtipo molecular de su tumor primario corresponde a Basal-Like. La paciente 27PNVV está, también, a la cabeza en el número de genes mutados y mutaciones, 7 y 10, respectivamente.

4.4.2.3. Alteraciones en T1 según la respuesta a la NAC

Aun siendo conscientes de que el subgrupo de pacientes que obtienen pCR es muy reducido, hemos dividido las descriptivas entre pacientes que obtienen pCR y los que no. Cabe destacar que tanto en pacientes que responden como en los que no, los resultados en nuestra cohorte son similares. El gen más mutado en ambos grupos es *TP53*, estando en el 45% de las pacientes que no responden al tratamiento y en el 100% de las pacientes que responden. El siguiente gen más mutado es *ATM*, tanto es respondedoras como en no respondedoras, en un porcentaje más similar que en el caso de *TP53*, 25% en pacientes no respondedoras y 33% en respondedoras. Los siguientes genes más mutados en las pacientes no respondedoras al tratamiento con NAC son *BRCA2*, *RECQL4*, estando en el 20 % de las pacientes con mutaciones. En el caso de las pacientes respondedoras, los genes *AURKA*, *CCND3*, *ESR1*, *KMT2C*, *NCOR1*, *PIK3CA* y *TBX3* se encuentran en el 33% de los casos. Estos últimos genes mencionados también se encuentran mutados en las pacientes no respondedoras, pero en menor porcentaje, excepto el gen *NCOR1* que no se encuentra mutado en las pacientes que no responde (Figura 29, B). En las pacientes que obtienen pCR vemos una homogeneidad en cuanto al tipo de mutación que encontramos, en este caso el 100% de las mutaciones son *missense*. Distinto es el caso de las que no obtienen pCR, aun siendo la más abundante la mutación *missense*, encontramos un abanico más amplio de mutaciones. En el caso de *TP53*, encontramos mutaciones tanto *missense* como en el sitio de ajustamiento o *splicing*, en el gen *ATM* se ha determinado tanto *missense* como *nonsense*, y en *BRCA2* y *RECQL4* hemos identificado *missense*, *nonsense* y alteración del marco de lectura por inserción (o *Frameshift insertion mutation*) en el caso de *BRCA2* y *missense* y *multi hit* en el caso de *RECQL4* (B y C de la Figura 29).

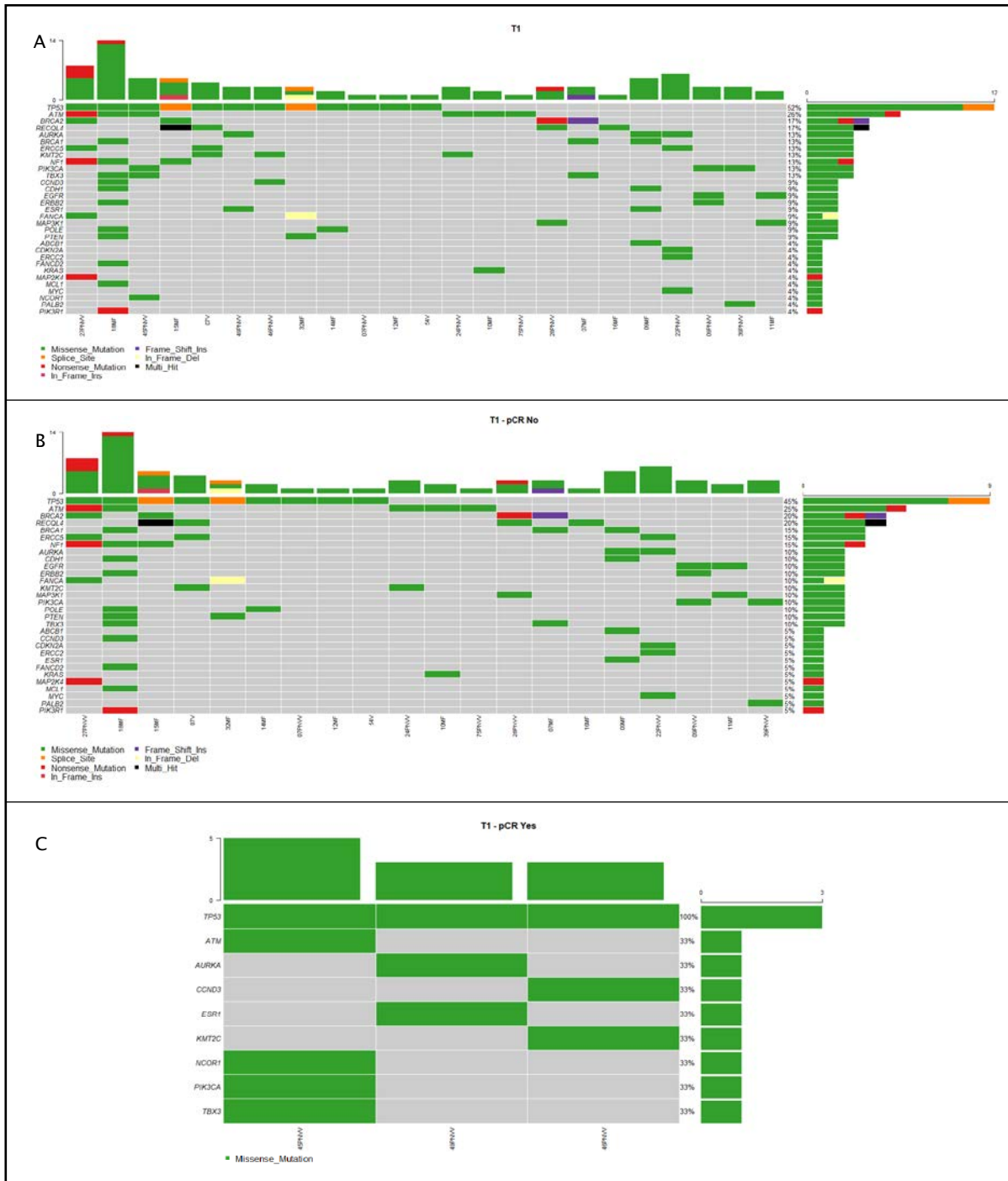


Figura 29. Oncoplots. Se identifican los genes en los que se han encontrado al menos una mutación y las pacientes de la DNA-cohorte. El porcentaje mostrado en la derecha de la matriz se corresponde con el porcentaje de mujeres con mutaciones que tienen ese gen concreto mutado. El diagrama de barras en el lateral derecho especifica el tipo de mutación encontrada dentro de cada gen. A. Resultados de todas las pacientes incluidas en el estudio. B. Resultados de las pacientes que no obtienen respuesta a la NAC. C. Resultados de las pacientes que obtienen respuesta a la NAC.

4.5. DISCUSIÓN

Identificar aquellos pacientes que tras la NAC van a conseguir pCR sería crucial para el abordaje de la enfermedad desde el inicio ya que sólo el 31% de las pacientes con cáncer de mama responden satisfactoriamente al tratamiento neoadyuvante. Actualmente, los mejores predictores de pCR se pueden dividir en dos clases, en primer lugar, aquellos que se identifican como variables clínico-patológicas determinadas en la mayoría de los casos por marcaje de proteínas, como es el caso de Ki67 o la determinación de los subtipos de mama por IHQ, por técnicas de imagen, como es el caso del tamaño tumoral mamográfico o definidas por la propia experiencia del facultativo como es el caso de la afectación ganglionar antes del tratamiento. En segundo lugar, se encuentran los predictores basados en las características moleculares como son la expresión génica o la determinación de alteraciones a nivel de genoma, como en el caso de los subtipos determinados por PAM50, la clasificación de *Inclust* o la firma TRAR (1). En base a toda esta información, el objetivo del estudio que hemos llevado a cabo es encontrar una firma que mejore lo ya conocido para obtener una herramienta predictora suficientemente robusta para la práctica clínica. En este trabajo, se han recogido todas las variables clinicopatológicas que por rutina se llevan a cabo en los hospitales que han colaborado en el proyecto (incluyen todas las variables definidas como predictoras de pCR en la literatura), se ha determinado el subtipo IHQ así como el subtipo de PAM50 en todas nuestras pacientes y además hemos analizado la expresión génica de 730 genes involucrados en cáncer y las mutaciones de los 52 genes más importantes en el cáncer de mama, todo ello con el objetivo de obtener un modelo que mejore la predicción de pCR. Como hemos mencionado anteriormente, el análisis de las mutaciones ha quedado reducido a una descriptiva de las alteraciones encontradas ya que el número de pacientes con DNA de calidad para realizar los experimentos se ha visto muy reducido. Cabe destacar que los genes mutados más frecuentemente en nuestras pacientes son *TP53* y *ATM*. A excepción del gen *NCOR1*, que se encuentra mutado exclusivamente en las pacientes con pCR, todos los genes mutados en las pacientes con pCR se encuentran también en las no respondedoras.

Sin embargo, debido al bajo número de pacientes incluidos en esta parte del estudio, estos resultados carecen de significación estadística y pueden presentar cierto sesgo. Por lo tanto, en la discusión nos centraremos en los resultados más prometedores generados en base al análisis de las diferencias en expresión génica entre pacientes respondedoras y resistentes a la NAC.

Uno de los principales hallazgos de este trabajo se basa en el hecho de que los modelos predictivos generados no son un subgrupo de los subtipos de cáncer de mama, incluso de aquellos subtipos determinados por variantes de expresión génica, sino que se basan en las diferencias biológicas encontradas en la expresión génica de 730 genes estrechamente relacionados con cáncer entre los dos escenarios de interés, tumores sin capacidad de hacer frente a la NAC, versus aquellos tumores con herramientas moleculares para hacer frente al tratamiento. De los 10 genes DE con significación estadística, 5 presentan un FC superior al 1.5 en valor absoluto (*FGFR4*, *CDC6*, *GNG4*, *MYCN* y *LEFTY1*), lo que indica que son los genes con mayor cantidad de cambio en su expresión génica entre pCR y no-pCR. A excepción de *LEFTY1* ($FC > -1.5$), dichos genes tienen mayor expresión en pacientes con pCR ($FC > 1.5$). Según el OR del análisis univariante, la expresión de los genes relacionada con pCR sigue la misma tendencia que el FC, es decir, en aquellos genes donde el FC es positivo, el OR está por encima de 1, y en el caso de FC negativo, el OR está por debajo de 1, reforzando estos resultados la validez de nuestros datos. Según los resultados obtenidos en el análisis de *Pearson*, las correlaciones en la expresión génica entre los pares de genes *FGFR4*-*MYCN* y *MYCN*-*GNG4* son las más intensas, positivas y estadísticamente significativas ($r=0.39$, $p=0.0002$ y $r=0.37$, $p=0.0005$, respectivamente). Ambos genes pertenecen a la vía de la PI3K, y están asociados positivamente a pCR; sin embargo, no existe correlación entre ellos, por lo que, nuestro análisis define una relación independiente con *MYCN* y por tanto la redundancia en cuanto a los mecanismos de resistencia asociados a esta ruta biológica y la necesidad de incluir varios genes en la descripción de estos mecanismos y de las posibles dianas terapéuticas.

Todos los modelos predictivos han sido diseñados en base a la importancia que tienen las diferentes variables a la hora de predecir respuesta al tratamiento. En nuestra cohorte

Allsubtypes, donde no hay distinción por subtipos, los modelos predictivos que incluyen los genes DE, como es el caso de los modelos *GENEX_Allsubtypes* y *CLINGENEX_Allsubtypes* obtienen mayor valor predictivo que aquellos que solo incluyen variables clínicas (*CLIN_Allsubtypes*) o el modelo formado únicamente por PAM50 (AUC de 0.79 y 0.74 vs 0.55 y 0.51, respectivamente). Cabe destacar que la variable con mayor importancia seleccionada en el modelo *CLIN_Allsubtypes* es la sobreexpresión de HER2, en coherencia con el hecho de que HER2 es uno de los marcadores más significativos de respuesta al tratamiento con NAC (4). En cuanto a la ratio entre nuestros modelos y el modelo PAM50, para determinar la cantidad de mejora con respecto a este clasificador que se postula como el mejor predictor de pCR en el panorama actual, vemos que tanto *CLINGENEX_ALLsubtypes* como *GENEX_Allsubtypes* son superiores en cuanto a poder predictivo definido por su valor de AUC. Esto se mantiene en las cohortes HER2- y HER2-/ER+. El valor más alto de la ratio se da en la Cohorte_D_HER2-, donde el modelo *CLINGENEX_HER2-* supera en más del doble al poder predictivo de PAM50. Resultados similares se obtienen con el modelo *GENEX_HER2-*. En la cohorte exclusivamente ER+ (HER2-/ER+) las ratios son menores. Como ya Prat y colaboradores (2016) identificaron, la capacidad predictiva de pCR del clasificador PAM50 en cohortes Luminales A (cohorte que clínicamente tiene un porcentaje alto de tumores ER+) era la única estadísticamente significativa (20). Sin embargo, aunque menor con respecto a la cohorte en general, el poder predictivo de *CLINGENEX_ER+* y *GENEX_ER+* sigue siendo mayor que el de PAM50. Sin atender a comparaciones con predictores ya publicados, cuando validamos nuestros modelos en un set de datos independiente al nuestro con las características de las cohortes donde fueron definidos, los resultados siguen la misma tendencia; por sí solas las variables clínicas no consiguen el mismo poder predictivo que cuando se integra la información de los genes DE de nuestra cohorte. El modelo *CLINGENEX_Allsubtypes*, aunque en la cohorte de descubrimiento presenta menor AUC que el modelo *GENEX_Allsubtypes*, en la cohorte de validación donde el número de pacientes aumenta considerablemente (84 vs 177), se posiciona como el mejor modelo predictivo. Aunque los resultados obtenidos en la cohorte_D_HER2-, con el modelo *CLINGENEX_HER2-* suponen una gran mejora con respecto a *CLIN_HER2-* o PAM50 (AUC 0.85 vs 0.60 y 0.52,

respectivamente) este modelo en la cohorte de validación pierde fuerza predictora (0.85 vs 0.67); si bien este modelo predice pCR en esta cohorte mejor que las variantes clínicas por separado. Y, por último, en aquellos tumores HER2-/ER+, la capacidad predictiva del modelo CLINGENEX_ER+ lo posiciona como el mejor modelo predictivo tanto en la cohorte_D_ER+, donde incluimos 39 pacientes, así como en la de validación, cohorte_V_ER+, donde se incluyen 344 pacientes (AUC 0.96 y 0.76, respetivamente). Teniendo en cuenta que las pacientes con tumores ER+ se benefician muy poco del tratamiento con NAC, tener una firma con un alto valor predictivo de pCR, como es el caso de CLINGENEX_ER+ (AUC 0.96 en la cohorte de descubrimiento y AUC 0.76 en la cohorte de validación) tendría alta utilidad clínica. Identificar los tumores resistentes a la NAC previamente a su administración posibilita la consideración de estrategias terapéuticas alternativas para este tipo de tumores como es la inmunoterapia o el tratamiento con inhibidores de CDK4/6, además de evitar los efectos tóxicos derivados de ella.

FGFR4 y *LEFTY1* son los genes que los modelos predictivos siempre eligen como variables con alta importancia para la predicción de pCR en todas las cohortes donde hemos diseñado modelos predictivos, es decir, en la cohorte que alberga todos los subtipos, en las que solo se incluyen pacientes HER2- y en la cohorte HER2-/ER+. *FGFR4*, que se encuentra más expresado es pacientes que obtienen pCR, es un miembro altamente conservado de la familia de los receptores tirosín-kinasa de membrana y es integrante de varias vías normalmente alteradas en procesos tumorales como son las vías de la MAPK, RAS y PI3K. El 7% de las alteraciones de los genes de la familia FGFR descritas en un estudio realizado en 4853 cánceres se encuentran en dicho gen (105). Y aunque haya sido históricamente poco estudiado en comparación con otros integrantes de la misma familia, la activación aberrante de *FGFR4* ha sido relacionada con progresión tumoral, supervivencia celular y resistencia a diferentes tratamientos. De hecho, cada vez hay más claras evidencias de que tiene un rol importante en el cáncer de mama y en el hepatocarcinoma. De este último ya hay diferentes ensayos clínicos en marcha (106). Según *The Cancer Genome Atlas* (TCGA), el gen *FGFR4* está más expresados en pacientes con cáncer de mama definidos como HER2enriched por PAM50 (22). Dado que el subtipo

HER2enriched es de los que mejor responden al tratamiento con NAC y mayor tasa de pCR obtiene (1), la sobreexpresión de *FGFR4* en nuestra cohorte de pacientes que tras el tratamiento obtienen pCR puede deberse a la presencia del subtipo HER2enriched, que es el 26% del total, y de ellas el 39% obtienen pCR. En un estudio reciente, García-Recio et al (2020) en modelos animales derivados de pacientes (PDX), relacionaron la sobreexpresión de *FGFR4* en el subtipo HER2-enriched previamente publicado en TCGA (22) con mayor progresión tumoral, metástasis y diferenciación de subtipos desde HER2enriched a HER2- tras el tratamiento con anti-HER2 (107). En cáncer de mama TN, se determinó que bajo un tratamiento específico donde no se incluía Antraciclinas, sino que eran tratados con Taxanos+Carboplatino o Taxano+Gemcitabina, únicamente en el caso del tratamiento con este último, la sobreexpresión del gen *FGFR4* se asociaba a una disminución de la tasa de pCR (11). Aunque a priori todo indique que la activación de *FGFR4* es un marcador de resistencia al tratamiento diana anti-HER2, y haya sido relacionado con una baja tasa de pCR en pacientes TN tratadas con Taxanos+Gemcitabina, en nuestra cohorte donde están presentes todos los subtipos y el tratamiento es de Antraciclinas+/-Taxanos en neoadyuvancia la sobreexpresión de dicho gen en la biopsia diagnóstica tras la NAC se relaciona con alta tasa de pCR (OR 1.48, 95% IC (1.20-1.92), $p=4.03E^{-05}$). Cabe destacar que esta implicación en la predicción de la pCR no la hemos visto únicamente en la cohorte *Allsubtypes*, sino que tanto en tumores HER2- como en HER2-/ER+ ha sido seleccionada como una de las variables con mayor peso para la predicción de la respuesta.

El otro gen que nos genera gran interés debido a su implicación en la predicción de pCR según nuestros datos y con independencia del subtipo es *LEFTY1*. Es un gen supresor de tumores, en trabajos donde se han estudiado masivamente el comportamiento de genes en el cáncer de mama, como es el caso del TCGA (22) o estudios procedentes del *Molecular Taxonomy of Breast Cancer International Consortium* (METABRIC)(108), se ha visto que el 20% de los cánceres de mama tienen amplificaciones del gen *LEFTY1*. Según los datos de Zabala et al. (2020), *LEFTY1* induce la proliferación en las células progenitoras del epitelio mamario mediante la inhibición dual de las vías SMAD2 y SMAD5, ya que la inhibición química de estas dos proteínas confiere mayor capacidad

proliferativa en las células epiteliales in vitro. Esto se relaciona con fenotipos *Stemness* (23). Además, el aumento de expresión de dicho gen ha sido asociado con capacidad proliferativa a largo plazo y esto a su vez con resistencia al tratamiento tanto con radiación como con quimioterapia (97). En nuestra cohorte *LEFTY1* tiene expresión a la baja en aquellos pacientes que responden a la NAC. EL hecho de que su expresión al alza se relacione en previos estudios con aumento de la proliferación no concuerda con el hecho de que el aumento de proliferación favorece la respuesta al tratamiento con NAC, pero su asociación con fenotipos *Stemness* y con capacidad proliferativa a largo plazo, lo posicionan como un posible marcador de *Cancer Stem Cell* y, por lo tanto, de resistencia a tratamiento. Por otra parte, se han descrito amplificaciones de *LEFTY1* en el 20% de los cánceres de mama, pero no hay ningún ensayo previo en el que se haya estudiado el comportamiento de este gen con respecto al tratamiento con NAC. Según nuestros resultados tanto del análisis de expresión diferencial como del análisis univariante, *LEFTY1* se asocia de manera negativa con la pCR, y en pacientes que responden a la NAC se encuentra más inhibida su expresión génica (análisis de expresión diferencial: FC -1.51 y p=0.00024, análisis univariante: OR 0.75 (95% IC 0.60–0.91) y p=0.0047).

4.6. CONCLUSIONES

En base a los resultados obtenidos en el estudio de nuestra cohorte, proponemos nuevas firmas génicas que predicen mejor la respuesta al tratamiento con NAC en pacientes con cáncer de mama. Dichos modelos predictivos, que incluyen las características de mayor importancia en la predicción de pCR integrando los marcadores clásicos, junto con la información de expresión génica y usando inteligencia artificial, definen mejor a la población que va a obtener pCR que las variables clínico-patológicas clásicamente usadas o que el clasificador de subtipos moleculares PAM50 por sí solo. Además, nuestros resultados, validados en un set de datos con un número de pacientes significativamente mayor a nuestra cohorte de estudio, muestran que aquellos tumores de mama que se han diagnosticados HER2-/ER+ al inicio, podrían beneficiarse del valor predictivo de la respuesta al tratamiento con NAC del modelo CLINGENEX_ER+, que según nuestros resultados mejoran al clasificador PAM50. Proponemos, también, que

los genes *FGFR4* y *LEFTY1* deben ser investigados por su implicación en los mecanismos de resistencia a NAC, al igual que los genes de los que ellos dependen o viceversa, ya que tienen un peso importante en la predicción de la respuesta a la NAC. En el caso del *FGFR4*, se encuentra muy asociado a responder satisfactoriamente a la NAC, , mientras que *LEFTY1* se asocia a mecanismos de resistencia, que podría deberse a su implicación en fenotipos *Stemness*.

5. Capítulo II:

Resistencia al tratamiento con NAC: Paisaje molecular de la enfermedad residual y la primera diseminación metastásica

Lo importante es no tener arrugas en el cerebro.

Margarita salas

5.1. INTRODUCCIÓN Y OBJETIVOS

La respuesta al tratamiento con NAC es clave para el pronóstico del paciente después de la cirugía. Existen distintos grados de respuesta a este tratamiento que van desde pCR, definido como ausencia células tumorales en el tejido mamario y en ganglios adyacentes, hasta una extensa enfermedad residual pasando por un todo un espectro de respuestas intermedias (1). La presencia de RD después del tratamiento con NAC es un factor pronóstico fundamental. Se ha demostrado que la reducción de la carga tumoral después del tratamiento está asociada con un incremento de SLE y de la SG (12) (3) (53) (13) . Con la idea de poder clasificar el resultado clínico según las características de la RD, se han determinado biomarcadores que proporcionan una valoración cuantitativa de la extensión tumoral y eso a su vez se ha relacionado con un resultado clínico concreto. La determinación de Ki67 en la RD, el índice de RCB y la RPCB son algunos de los marcadores que se usan para determinar los distintos espectros de resistencia (1).

Cuando el porcentaje de Ki67 es bajo tras el tratamiento (0–15%) el resultado clínico es comparable a la ausencia de enfermedad residual u obtención de pCR. Pacientes con RD con un porcentaje alto de Ki67 (>35%) tienen más riesgo de recaída y de muerte que el resto de las pacientes con porcentajes menores (57) (66).

El índice RCB estratifica los pacientes con RD tras la NAC según una serie de parámetros anatomo-patológicos del tumor residual que se corresponden con un resultado clínico concreto; RCB-0 (equivalente a pCR), RCB-I (RD mínima), RCB-II (RD moderada) y RCB-III (RD extensa). De hecho, este índice ha sido identificado por Symmans y colaboradores como un predictor independiente de la SLRD (49). Una sinergia entre los dos marcadores anteriores es el RPCB, que aglutina el valor de Ki67 y el nivel de RCB de la RD, resulta mucho más útil en el escenario del pronóstico a largo plazo. Es más, este marcador es de valor pronóstico incluso en pacientes con tumores ER+, en los que el valor Ki67 y RCB por separado carece de él (69).

5.1.1. Marcadores moleculares de RD

Debido a la agresividad que presenta el cáncer de mama TN cuando el efecto citotóxico de la NAC no acaba con la totalidad del tumor, es el tipo de cáncer de mama donde más ha sido estudiada la RD a nivel molecular. De hecho, en este subtipo se ha comprobado que la NAC trae consigo una serie de adquisiciones y pérdidas caóticas de CNAs en el genoma de las células que sobreviven a ella. Por ejemplo, las amplificaciones en los genes *MCL1*, *MYC*, *KRAS*, *BCL9*, *EGFR*, *CDKN2A*, *BIRC2* y *BIRC3* han sido identificadas en la RD cuando se compara con el tumor antes del tratamiento (24). Además, perfiles específicos de expresión génica del tumor residual también tienen valor pronóstico. De hecho, en pacientes TN, se han generado firmas de expresión que forman un modelo predictivo para la SLR en tres años con los genes *AR*, *GATA3*, *ESR2*, *GBX2*, *KRT16*, *MMP28* y *WNT11*. Aquellas pacientes TN que en la RD expresen al alza genes como *AR* o *GATA3* (genes relacionados con fenotipos Luminales) van a tener una SLR del 76.9% frente al 25% de las pacientes que expresen al alza genes relacionados con marcadores de CSCs como es el caso de *KRT16*, *WNT11* o *MMP28* (109).

5.1.2. Dinamismo en los biomarcadores ER, PR y HER2 tras la NAC

Debido a esta serie de cambios refractarios al tratamiento con NAC en la biología molecular del tumor de mama, se establece que este tumor residual adquiere una identidad tumoral propia, distinta al tumor primario, variable y que como hemos indicado puede tener valor pronóstico. Entre esos cambios moleculares definitorios de fenotipos específicos dentro de la enfermedad residual se encuentra la mutabilidad en la expresión de los marcadores por excelencia (además de dianas terapéuticas) en el cáncer de mama, como son los receptores hormonales ER, y de progesterona PR y HER2, determinada en estudios donde se comparaban las similitudes/diferencias entre el tumor primario y la RD. En el caso de los receptores hormonales, en tumores residuales suele observarse la negativización de estos durante el tratamiento. En base a esto, una positividad en el tumor tras el tratamiento puede indicar la presencia de alta heterogeneidad en el tumor primario. En el caso de los tumores con sobreexpresión del HER2 tratados con NAC y Trastuzumab, el 43% de las pacientes pierden en la RD la

sobreexpresión de dicho marcador, debido, sobre todo, al tratamiento diana anti-HER2 con el que son tratados además del tratamiento citotóxico. Esto conlleva una eliminación específica de aquellas células que sobreexpresan HER2.

Las consecuencias pronósticas de estas pérdidas de dianas terapéuticas en la enfermedad residual son claras (57) (71) (70). Por tanto, la caracterización no sólo de la extensión tumoral con los marcadores tradicionales, sino del fenotipo concreto del tumor residual de cada paciente, es imprescindible para la valoración pronóstica de la enfermedad residual y la elección del siguiente tratamiento.

Más allá de la enfermedad residual, se carece en gran medida de factores deterministas de metástasis después del tratamiento con NAC y de un conocimiento más profundo de su relación con perfiles moleculares específicos del tumor primario y la RD que puedan generar modelos predictivos de la enfermedad a distancia en los distintos momentos naturales de la enfermedad y el tratamiento. Este aspecto es sumamente importante si consideramos que la enfermedad a distancia o metástasis es responsable del 90% de las muertes relacionadas con el cáncer de mama (25). Actualmente, hay estudios donde se han establecido varios marcadores diagnósticos de cáncer de mama metastásico; sin embargo, no se ha determinado ningún predictor único de metástasis a partir de la RD. Considerando la necesidad expuesta de identificación de marcadores pronósticos y que definan el fenotipo de enfermedad residual, así como predictores de metástasis desde los distintos momentos de la enfermedad y el tratamiento, definimos los **objetivos** en este capítulo como:

- 1) Identificación de las alteraciones moleculares entre el tumor primario y la enfermedad residual para entender la evolución tumoral e identificar marcadores específicos de resistencia.
- 2) Comparación de las alteraciones encontradas entre el tumor primario y la enfermedad residual con respecto al primer nicho metastásico para definir una posible respuesta inmutable desde el inicio de la enfermedad.

Para ello, estudiaremos la expresión génica de 730 genes íntimamente relacionados con cáncer con un análisis longitudinal de **una cohorte homogénea de pacientes con cáncer de mama que presentaron enfermedad residual con posterior metástasis a distancia,**

serán excluidas todas aquellas pacientes que presenten metástasis antes del tratamiento con NAC. En todos los casos en los que ha sido posible, el estudio incluye la muestra del tumor primario o muestra pre-tratamiento (**Pre_T1**), la enfermedad residual (**RD_T2**) y la metástasis (**Mtx_T3**).

5.2. POBLACIÓN DE ESTUDIO Y METODOLOGÍA

A continuación, se va a describir los pacientes y la metodología específica de este capítulo.

5.2.1. Pacientes y muestras

Se ha llevado a cabo un análisis retrospectivo con pacientes bajo los siguientes criterios de inclusión: que sean mujeres adultas (mayores de 18 años) con cáncer de mama tratadas con NAC (Antraciclinas+/-Taxanos y terapia dirigida con Trastuzumab en las pacientes HER2+) y **no respondedoras** al tratamiento con enfermedad residual y metástasis. Las pacientes con subtipos IHQ determinados como TN y HER2+ no han recibido terapia hormonal tras la cirugía, mientras que aquellas determinadas como Luminales sí han recibido. Las pacientes que presenten metástasis en el momento del diagnóstico del tumor primario o antes del tratamiento con NAC serán descartadas. Que tuvieran, al menos, muestra de tejido tumoral tanto del tumor primario antes del tratamiento (**Pre_T1**) como de la enfermedad residual (**RD_T2**) y en los casos donde sea posible muestra del tejido tumoral del primer nicho metastásico (**Mtx_T3**).

Todos los análisis se han llevado a cabo en muestras de tejido incluidos en parafina. La respuesta al tratamiento se ha definido en la biopsia de la cirugía tras el tratamiento neoadyuvante, según el criterio Miller & Payne. La distribución de las pacientes es muestra en la figura 30.

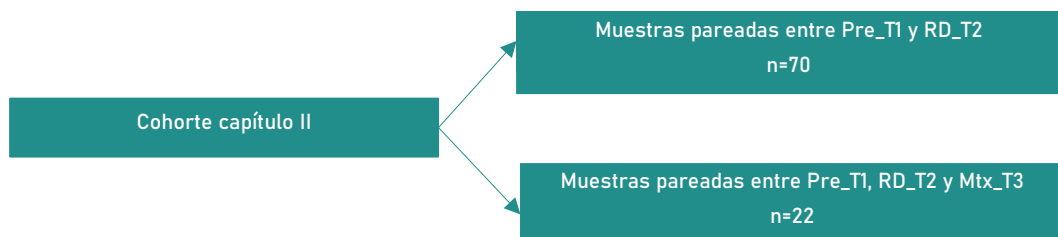


Figura 30. Esquema de la cohorte del capítulo II.

5.2.2. Inmunohistoquímica, subtipos por PAM50 y determinación de la RD

De cara a conocer nuestra población de estudio desde una perspectiva clínica y para dividir las pacientes en los diferentes subtipos de cáncer de mama, se realizaron tinciones inmunohistoquímicas para determinar ER, PR, CK5/6, Ki67 y HER2.

Desde los bloques de parafina de los pacientes seleccionados para el presente estudio se hicieron cortes de 5µm. La estación de tinción automatizada (Autostainer Plus, Dako) fue usada para ello, usando el kit EnVision FLEX System (Dako). Como cromógeno, usamos diambencidina y para reemplazar el anticuerpo (Ac) primario se usó un suero no-inmune como control negativo, y un control positivo específico para cada anticuerpo.

A continuación se describen los anticuerpos que se usaron para la cuantificación de las proteínas: para el receptor de estrógenos (ER) se usó el Ac monoclonal de conejo, clona SP1 de Master diagnostica; para el receptor de progesterona (PR), el Ac monoclonal de conejo, clona Y85 de Master diagnostica; para la citoqueratina 5/6 (CK5/6), el Ac monoclonal de ratón, clona D5/16B4 de Master Diagnostica; para el EGFR, el Ac monoclonal de conejo, clona EP38Y de Master Diagnostica; para cuantificar Ki-67, usamos el Ac monoclonal de conejo, clona SP6 de Master Diagnostica; y para HER2 el Kit Herceptest de Dako. La tinción fue evaluada por el grupo de anatomía patológica. Para ER y PR el resultado positivo se define como toda aquella tinción nuclear en más del 1% de las células (83) y para EGFR y CK5/6, la positividad viene dada por la detección membranosa o citoplasmática en al menos una célula. El marcador de proliferación Ki67 fue determinado cuantificando el porcentaje de células neoplásicas con tinción nuclear de Ki67. La sobreexpresión de HER2 fue evaluada de acuerdo a las guías internacionales (84) y categorizadas como negativas si el resultado de la inmunohistoquímica era 0 o +1; cuando el resultado era +2, se procedió a determinar la amplificación del gen por hibridación *in situ* fluorescente (FISH) o por hibridación *in situ* cromogénica (CISH).

Los subtipos moleculares fueron determinados con el clasificador PAM50 de Nanostring Technologies, basado en la medición de la expresión de 50 genes que clasifican los tumores en 4 subtipos, Luminal A, Luminal B, HER2-enriched y Basal-Like (43).

Para la determinación de la RD, con una tinción de Hematoxilina y Eosina, el patólogo determinó un valor dentro de la clasificación de Miller and Payne, donde el 5 equivale a ausencia de células tumorales y del 4-1 equivale a diferentes grados de presencia de células tumorales. Es decir, aquellas pacientes con valores entre 4 y 1 en la clasificación de Miller and Payne (en la muestra de la mama o en la axila) se consideran pacientes con tumores resistentes al tratamiento con NAC o, dicho de otra manera, pacientes con RD, mientras que las pacientes que obtienen un score de 5 presentan pCR.

5.2.3. Determinación de la expresión génica

5.2.3.1. Panel PanCancer Pathways

Según los objetivos que hemos especificado con anterioridad, se ha determinado la expresión de 730 genes, más 40 genes *housekeeping*, estrechamente relacionados con cáncer en las 163 muestras (T1, T2 y T3) incluidas en esta parte del estudio. La tecnología para determinar la expresión génica ha sido *nCounter*.

Se ha llevado a cabo un **AED** con la herramienta nSolver™, donde el p-valor fue ajustado por el método *Benjamini-Hochberg*. El límite estadístico para definir a un gen DE entre cada grupo comparativo se establece en un p-valor de 0.05, de manera que aquellos genes cuya expresión sea diferente con una significación estadística por debajo de 0.05 serán considerados DE, independientemente de la magnitud de la diferencia o Log2 FC.

5.3. RESULTADOS

5.3.1. Características clínico-patológicas

Las pacientes incluidas en este estudio han sido divididas en dos, por un lado, aquellas pacientes que solo tienen muestra del tumor primario y la RD (n=70) y por otro lado aquellas pacientes con muestras del tumor primario, RD y primera localización metastásica (n=22). De las pacientes que han sido estudiadas en la cohorte Pre_T1vsRD_T2 el 68% eran premenopáusicas, tenían un tamaño tumoral de 4 cm de media, en el 57% de los casos tengo los ganglios linfáticos afectados, el 55% tenían grado histológico entre 1 y 2, el 61% son ER+, el 52% PR+ y el 17% HER2+. El 58% tiene menos

del 50% de Ki67, y los subtipos intrínsecos están divididos en 9 Luminal A, 26 Luminal B, 14 HER2enriched y 19 Basal-Like (Tabla 19).

En la cohorte de pacientes donde hemos estudiado la relación del tumor primario, la enfermedad residual y la localización metastásica Pre_T1/RD_T2/Mtx_T3, el 54% eran premenopáusicas, el tamaño tumoral estaba en 3.7 cm de media, el 54% tienen ganglios linfáticos afectados, el 60% tiene tumores con grado histológico entre 1 y 2, el 50% son ER+, el 45% PR+ y el 27% HER2+. El 68% de las pacientes tienen menos del 50% de Ki67, hay 5 Luminales A, 4 Luminales B, 6 HER2enriched y 5 Basal-Like determinado por PAM50 y la localización de las metástasis se distribuye en el 27.27% en hueso, el 22.73% en cerebro, el 22.73% en piel (región mamaria), el 13.64% en hígado, el 4.55% en Ovario y el 9.09% en ganglios linfáticos (Tabla 20).

Tabla 19. Características clínico-patológicas de las pacientes incluidas en la cohorte Pre_T1 vs RD_T2.

Variables clinicopatológicas	Pre_T1 vs RD_T2	
	Eventos	%
Todos los casos	70	
Edad	48	[30 - 75]
Estatus menopáusico		
Premenopáusico	48	68.57
postmenopáusico	22	31.43
Tamaño tumoral	4	[0,97 - 12,00]
Afectación ganglios linfáticos		
Positivos	40	57.14
Negativos	28	40.00
NA	2	2.86
Grado histológico tumoral		
1 y 2	39	55.71
3	24	34.29
NA	7	10.00
Estatus ER		
Positivo	43	61.43
Negativo	27	38.57
Estatus PR		
Positivo	37	52.86
Negativo	33	47.14
Sobreexpresión del HER2		
Positiva	12	17.14
Negativa	58	82.86
Porcentaje de Ki67		
Menos de 50	41	58.57
Más de 50	25	35.71
NA	4	5.71
Subtipos IHQ		
Luminal A	8	11.43
Luminal B	28	40.00
HER2+	6	8.57
Triple Negativo	19	27.14
NA	1	1.43
Subtipos PAM50		
Luminal A	9	12.86
Luminal B	26	37.14
HER2-enriched	14	20.00
Basal-Like	19	27.14
NA	2	2.86

Tabla 20. Características clínico-patológicas de las pacientes incluidas en la cohorte Pre_T1/RD_T2/Mtx_T3

Variables clinicopatológicas	Pre_T1/ RD_T2/Mtx_T3	
	Eventos	%
Todos los casos	22	
Edad	53	[30 – 73]
Estatus menopáusico		
Premenopáusico	12	54.55
postmenopáusico	10	45.45
Tamaño tumoral	3.7	[1 – 7.70]
Afectación ganglios linfáticos		
Positivos	12	54.55
Negativos	8	36.36
NA	2	9.09
Grado histológico tumoral		
1 y 2	13	59.09
3	8	36.36
NA	1	4.55
Estatus ER		
Positivo	11	50
Negativo	11	50
Estatus PR		
Positivo	10	45.5
Negativo	12	54.5
Sobreexpresión del HER2		
Positiva	6	27.27
Negativa	16	72.73
Porcentaje de Ki67		
Menos de 50	15	68.18
Más de 50	5	22.73
NA	2	9.09
Subtipos IHQ		
Luminal A	5	11.43
Luminal B	4	40
HER2+	5	8.57
Triple Negativo	6	27.14
NA	1	1.43
Subtipos PAM50		
Luminal A	5	12.86
Luminal B	4	37.14
HER2-enriched	6	20
Basal-Like	5	27.14
NA	2	2.86
Localización Mtx		
Hueso	6	27.27
Cerebro	5	22.73
Piel	5	22.73
Hígado	3	13.64
Ovario	1	4.55
Ganglios linfáticos	2	9.09

5.3.2. AED del panel *PanCancer Pathways* en Pre_T1, RD_T2 y Mtx_T3.

5.3.2.1. Tumor primario (Pre_T1) versus enfermedad residual (RD_T2).

Los genes DE entre aquellos pacientes con tumores primarios (Pre_T1) y enfermedad residual (RD_T2) tras el tratamiento con NAC están descritos en el *volcano plot* (Figura 31) y en la tabla 21. De los 730 genes que han sido estudiados, 112 genes se encuentran DE entre Pre_T1 y RD_T2, siendo la referencia el Pre_T1 (tabla 21). Vamos a entrar en detalle con los 10 genes que obtienen el p-valor más bajo, en este caso *FOS*, *NR4A1*, *NR4A3*, *CACNG4*, *IL6*, *COL11A1*, *TSPAN7*, *FLNC*, *ITGB6* y *KLF4*. En base al p-valor, el gen con mayor poder estadístico cuando hablamos de diferencia de expresión génica entre el Pre_T1 y RD_T2 es el gen *FOS* ($p=2.57E^{-23}$). Además, es el gen con más cantidad de cambio en su expresión génica, según su Log2 FC. Dicho gen está menos expresado en el Pre_T1 con respecto al RD_T2 (Log2 FC -3.55). Un aumento en la expresión génica de *FOS*, según nuestros datos, sería característico de aquellos tumores que no responden al tratamiento con NAC y por lo tanto la expresión de este gen podría estar asociado a resistencia a quimioterapia. *FOS* es un protooncogén, y junto a *JUN*, que en nuestra cohorte también se encuentra DE en RD ($p= 2.0E^{-03}$, Log2 FC de -1.1) forma el complejo del factor de transcripción AP-1. Está implicado en procesos de regulación de la proliferación celular, diferenciación y transformación. En determinados casos la expresión de *FOS* ha sido asociada a muerte celular apoptótica (RefSeq, 2008). Shi y colaboradores establecieron que la expresión de *FOS* podría estar relacionada con adaptación celular a cambios en el ambiente tumoral; estos resultados fueron obtenidos del estudio de la línea celular de mama MCF7. La baja expresión de *FOS* se ha asociado a una sensibilización celular a los agentes quimioterapéuticos en MCF7 tanto por la vía de la apoptosis como por la vía de la P-glicoproteína, es decir, que la expresión de *FOS* se relacionó con resistencia a la apoptosis, proponiéndose como un potencial marcador de resistencia a terapia (28). En nuestros resultados dicho gen se encuentra más expresado en aquellas pacientes resistentes a la quimioterapia Neoadyuvante, coherentemente con los hallazgos comentados.

Otros genes DE y que también se encuentran sobreexpresados en las pacientes con enfermedad residual y expresados a la baja en el tumor primario de nuestra población

son *NR4A1* y *NR4A3* ($p=5.4E^{-16}$ y Log2 FC de -2.8 para *NR4A1*; $p=4.4E^{-11}$ y Log2 FC de -2.2 para *NR4A3*). Ambos genes forman parte de la superfamilia de receptores de hormonas esteroideas-tiroideas-retinoides. En el caso de *NR4A1*, su expresión es inducida en linfocitos humanos y se relaciona con infiltración inmunológica tumoral. Codifica para una proteína que actúa como un factor de transcripción nuclear (RefSeq, 2011). El papel de *NR4A1* en el proceso tumoral no es claro y, además, controvertido, ya que en algunos estudios se le ha identificado como supresor de tumores y en otros como oncogén. En algunos de estos estudios se le relaciona con aumento de proliferación, viabilidad celular y apoptosis, pero sobre todo en roles de migración y progresión. En pacientes y líneas celulares TN, se ha determinado que existe una disminución de la expresión de esta proteína y ha sido relacionado con una atenuación de la vía JNK-AP-1-Ciclina D1 (29). *FOS*, en nuestra cohorte, está DE en la misma dirección que *NR4A1*, y como hemos comentado con anterioridad, forma parte del complejo AP-1 junto con JUN, lo que refuerza los resultados. Receptores de su misma naturaleza como el ER o el receptor de andrógenos (AR) se conoce muy bien su mecanismo de acción, pero en el caso de *NR4A1* no se le ha reconocido ligando específico, lo que lo descarta, por ahora, para tratamiento diana (29). En cáncer de mama, las citoquinas inflamatorias inducen la expresión de *NR4A1*; se ha comprobado que es un activador muy potente de la vía TGF- β , fosforilando a SMAD2, y contribuyendo a funciones antiapoptóticas, progresión tumoral, EMT y migración celular (30) (Figura tal y tal). Por otra parte, *NR4A3* es un activador transcripcional (RefSeq, 2010). Fedorova y colaboradores, en un estudio con tumores de mama y de pulmón, identificaron que las funciones supresoras de tumores de *NR4A3* se rigen por dos mecanismos distintos; el primero es dependiente de p53, pues es una diana transcripcional directa del mismo; el segundo consiste en que la sobreexpresión de *NR4A3* atenúa la proliferación de células cancerígenas y promueve la apoptosis a través de los genes pro-apoptóticos (Bcl-2), PUMA y Bax (31). Que este gen esté directamente relacionado con p53 y a su vez con atenuación de la proliferación, es coherente con su sobreexpresión en las pacientes resistentes a la NAC, ya que la activación de mecanismos de proliferación está estrechamente ligada a la respuesta a dicho tratamiento (1). Por otra parte, la función

anti-apoptótica asociada a la expresión de NR4A1 y la función pre-apoptótica asociada a la de NR4A3, parecen ser seleccionadas negativamente tras el tratamiento con NAC, debido a que ambos genes presentan la misma tendencia de expresión en nuestra cohorte, más expresados en la RD.

Una disminución de expresión de *CACNG4* es característico de tumores resistentes al tratamiento con NAC en nuestra cohorte ($p=1.1E-07$, Log2 FC 2). Es un gen que codifica para una proteína reguladora del receptor de AMPA transmembrana tipo I (TARP, *transmembrane AMPA receptor regulatory protein*), la subunidad Gamma 4 del canal de calcio regulado por voltaje. Los TARP regulan tanto el tráfico como la activación de canales de los receptores AMPA (RefSeq, 2010). *In vitro* se ha visto que la amplificación de este gen está relacionada con un incremento en la capacidad metastásica pulmonar en cáncer de mama, incrementando la supervivencia y la movilidad celular. Se sugiere que una vía subyacente al crecimiento y diseminación tumoral podría estar regulada por CACNG4, de manera que este cierra el canal del poro e inhibe la afluencia de calcio, alterando así los eventos de señalización con calcio que involucran genes clave de supervivencia y metástasis como el AKT2, HDAC3 o RASA1 (110). En células de adenocarcinoma de pulmón se ha visto que la expresión de este gen es mayor en las células sensibles al tratamiento con cisplatino que en células resistentes a dicho tratamiento (111), por lo que podemos hipotetizar que el bloqueo de dicha señalización también es importante para la eficacia a la NAC

EL gen *IL6* codifica para la interleucina 6, una citoquina que participa en el proceso inflamatorio y en la maduración de las células B (RefSeq, 2020). En un estudio de Dethlefsen y colaboradores se identificó una discrepancia en cuanto a la función de IL6 en los tumores de mama. En cultivo celulares de cáncer de mama ER+, una aplicación directa de IL6 inhibe la proliferación, mientras que altos niveles de IL6 circulante era correlacionado con mal pronóstico. Concluyeron el estudio con el carácter multifuncional de IL6, el elevado nivel sistémico es un biomarcador de carga tumoral, mientras que la expresión local intratumoral tiene un papel importante en el control del crecimiento celular, metástasis y la autorenovación de las *Cancer Stem Cells* (CSCs) (112). Con respecto a la formación de CSCs, a través del estudio de microRNAs en

cultivos celulares de cáncer de mama y en tumores de mama humanos, Iliopoulos y colaboradores, demostraron que IL6 era suficiente para que una célula no-CSC fuera convertida en CSC durante el proceso de transformación celular (113). Estas funciones intratumorales que describen Dethlefsen y col. y los resultados de Iliopoulos y col. de la IL6 están íntimamente relacionadas con mecanismos de resistencia al tratamiento citotóxico como es el caso de la NAC. En cultivos celulares de cáncer de mama HER2+, se demostró que la inflamación desarrollada por IL6 era mediador de la resistencia al tratamiento con Trastuzumab (114). En nuestra cohorte su expresión está relacionada con resistencia al tratamiento, estando más expresada en pacientes con RD tras la NAC ($p=2.3E^{-06}$, Log2 FC -1.8).

COL11A1 codifica para una de las dos cadenas alfa del colágeno tipo XI (RefSeq, 2009). La expresión de este gen ha sido relacionada en varios estudios con resistencia a quimioterapia en carcinoma epitelial de ovario, en unos casos por la activación de TWIST1 dando lugar a procesos de inhibición de la apoptosis (115), y en otros casos por la activación de la vía Akt/c/EBP β y ubiquitinación de *PDK1* (116). Además, también en cáncer de ovario se ha relacionado la expresión de este gen con progresión tumoral y un mal resultado clínico por la previa activación de la vía TGF- β 1 que desencadena la activación de la cascada de SMAD2 y a su vez la activación de COL11A1 (117). Esto en contrario a los resultados que obtenemos en nuestra cohorte de pacientes con cáncer de mama, donde la sobreexpresión de este gen se observa en el tumor primario con respecto a la RD, lo que indica carácter de sensibilidad a la NAC ($p=2.3E^{-06}$, Log FC 1.9). El gen **TSPAN7** codifica para un miembro de la superfamilia de transmembranas 4, con 4 dominios hidrofóbicos media la traducción de señales que tienen un papel en la regulación del desarrollo celular, activación, crecimiento y movilidad (RefSeq, 2008). En un estudio reciente (2021) la expresión a la baja de este gen ha sido asociada con un mal pronóstico es cáncer de vejiga, relacionan la expresión de este gen con activación de apoptosis y control del ciclo celular a través de la vía PTEN/PI3K/AKT(118). Sin embargo, no hay estudios concluyentes donde se relacione TSPAN7 con cáncer de mama. En nuestra cohorte, la expresión de este gen está relacionada con resistencia al

tratamiento con NAC, ya que se encuentra más expresados en la RD que en el tumor primario ($p=7E^{-06}$, $\text{Log}_2 \text{FC} -1.4$).

La **FLNC** codifica para la gamma filamina C, siendo la unión entre filamentos de actina. Está relacionada con las vías de la señalización de la MAPK y SMAD (RefSeq, 2008). Su función es esencial para la locomoción celular y migración en respuesta a un estímulo microambiental, estando más conservado en músculos cardiacos (119). En un estudio con pacientes con cáncer de mama tipo basales donde se comparaba el DNA en sangre, el tumor primario, la metástasis y modelos animales PDX, determinaron dos nuevas mutaciones en la muestra metastásica siendo una de ellas una mutación silenciosa en FLNC, concluyeron que esta mutación no era determinante en el proceso metastásico (120). Según nuestros resultados, la expresión de este gen está asociada a resistencia a tratamiento ($p=3.7E^{-05}$, $\text{Log}_2 \text{FC} -1.4$).

ITGB6 codifica para una proteína miembro de la superfamilia de las integrinas. Son receptores de adhesión que participan en la señalización desde la matriz celular a la célula (RefSeq, 2013). Tanto la expresión al alza del RNA mensajero de dicho gen como un aumento de la proteína han sido relacionados con una reducción en la supervivencia y aumento de la metástasis, sobre todo en pacientes que tienen sobreexpresados el HER2, posicionándose como una posible diana terapéutica para pacientes con cáncer de mama HER2+ con resistencia a Trastuzumab (121). Según nuestros resultados, una expresión al alza de este gen se relaciona con sensibilidad al tratamiento con quimioterapia neoadyuvante sin hacer distinciones con entre las pacientes que sobreexpresen HER2 y las que no ($p=8.8E^{-05}$, $\text{Log}_2 \text{FC} 1.5$).

Para un componente de la familia de factores de transcripción Kruppel codifica el gen **KLF4**, siendo requerido para el desarrollo normal de la función barrera de la piel. Se piensa que forma parte del control de la transición de G1 a S del ciclo celular después de presentar daños en el DNA al mediar el gen supresores de tumores p53 (RefSeq, 2015). Se le reconoce tanto como un oncogén, ya que se encuentra altamente expresados en más del 70% de los canceres de mama (122) y como un supresor de tumores por su participación en la red molecular de p53 (123). Se ha encontrado un aumento de expresión de este gen en las CSCs en líneas celulares de cáncer de mama y

en tumores primarios de mama en ratones (122). Junto con NANOG, se consideran factores inductores de células madre (*iPS*, *induced pluripotent stem*) implicados en el pronóstico de cáncer de mama; en subtipos TN se demostró que un aumento de la expresión de *KLF4* estaba relacionado con mejores ratios de SG y SLE, debido a su capacidad de inhibir la proliferación y la invasión en este subtipo de tumores de mama (124). Por otra parte, se ha demostrado que la activación de p53 disminuye la actividad transcripcional de ER por el aumento en la expresión de *KLF4*. *KLF4* inhibe la unión de ER α a los elementos respondedores del estrógeno en la región promotora, resultando en una reducción de la transcripción de los genes diana de ER α . Por lo que se descubrió una nueva red molecular entre p53, *KLF4* y ER. Tanto p53 como ER α están involucrados en crecimiento celular y apoptosis, es por ello que a *KLF4* se le atribuya tanto roles de supresor de tumores como funciones oncogénicas en cáncer de mama (123).

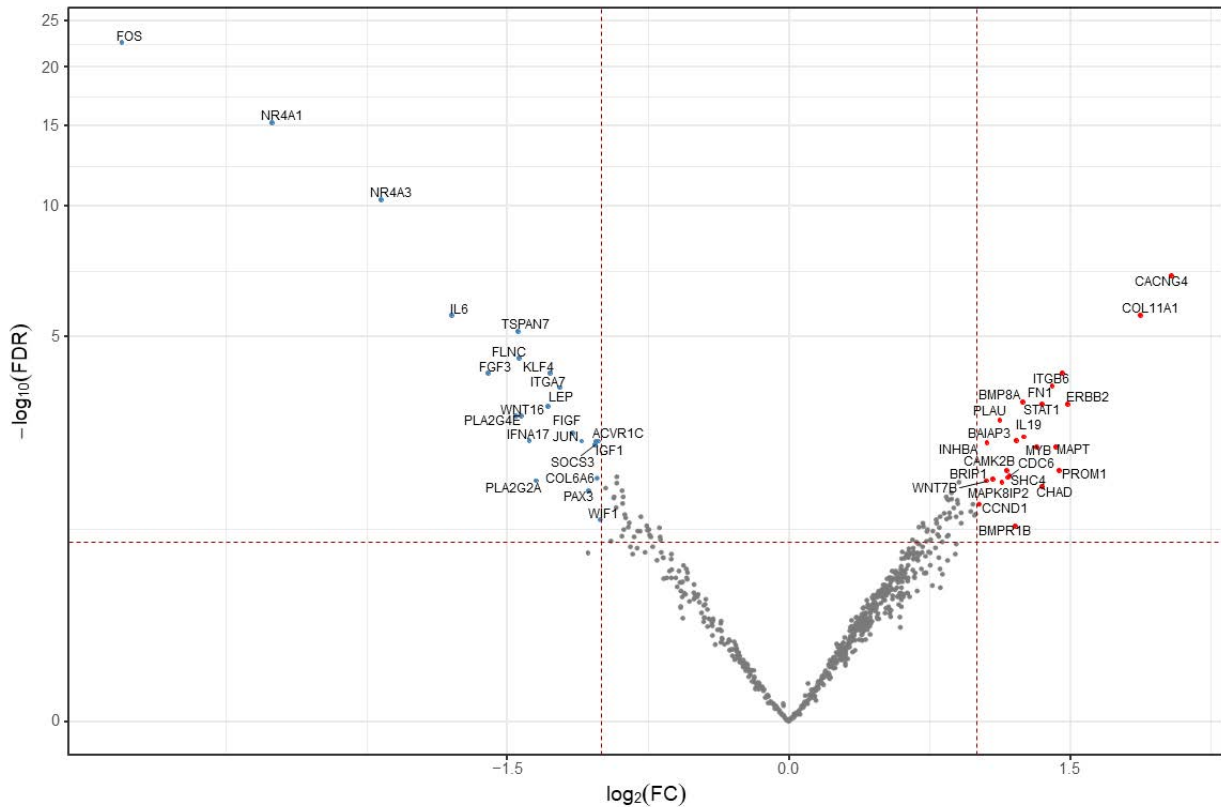


Figura 31. En el *volcano plot* se representan todos los genes según el p-valor ajustado obtenido en el AED (eje de ordenadas) y el valor del Log2 FC (eje de abscisas). Con las líneas punteadas se representan diferentes niveles de p-valor y la zona del *volcano* donde se posiciona el gen corresponderá con un p-valor concreto y un valor de FC determinado. En este caso, los genes que queden a la derecha del punto 0 del eje de abscisas serán los genes sobreexpresados en el Pre_T1 y los que queden en el lado izquierdo serán los sobreexpresados en el RD_T2.

Tabla 21. Genes DE entre Pre_T1 y RD_T2.

Gen	Log2 FC	p-valor ajustado	Gen	Log2 FC	p-valor ajustado	Gen	Log2 FC	p-valor ajustado
FOS	-3.6	2.6E-23	WNT7B	1.1	0.009	PRL	-0.9	0.031
NR4A1	-2.8	5.4E-16	PLA2G2A	-1.3	0.009	PLA2G10	0.9	0.031
NR4A3	-2.2	4.4E-11	MAPK8IP2	1.1	0.009	ITGA2	0.8	0.031
CACNG4	2.0	1.1E-07	WNT2	0.9	0.009	PDGFC	0.7	0.031
IL6	-1.8	2.3E-06	LEPR	-0.9	0.009	CSF3R	0.8	0.031
COL11A1	1.9	2.3E-06	CHAD	1.3	0.010	IL1R2	-0.8	0.033
TSPAN7	-1.4	7.0E-06	BMP2	-1.0	0.011	BMPR1B	1.2	0.033
FLNC	-1.4	3.7E-05	PAX3	-1.1	0.012	FGF1	0.7	0.034
ITGB6	1.5	8.8E-05	DUSP2	-0.9	0.012	SOCS2	-0.9	0.034
KLF4	-1.3	8.8E-05	RUNX1	0.9	0.013	SPRY4	0.7	0.034
FGF3	-1.6	8.8E-05	TLR2	0.8	0.014	ITGB4	0.8	0.035
FN1	1.4	1.7E-04	NGFR	-0.9	0.014	PCK1	-0.7	0.035
ITGA7	-1.2	1.8E-04	IL12B	-0.9	0.014	ETV1	0.7	0.037
BMP8A	1.2	3.7E-04	CACNA2D3	-0.9	0.015	GRIN1	-0.9	0.037
STAT1	1.3	4.1E-04	CACNB3	0.8	0.015	CACNA1G	-0.7	0.037
ERBB2	1.5	4.2E-04	KMT2C	0.9	0.015	TSLP	-0.8	0.037
LEP	-1.3	4.5E-04	MLLT4	0.9	0.016	PLA2G5	-0.8	0.038
WNT16	-1.4	7.0E-04	CDKN2B	0.9	0.016	H2AFX	0.8	0.038
PLA2G4E	-1.5	7.0E-04	TPO	-0.9	0.018	WNT11	-0.8	0.038
PLAU	1.1	8.7E-04	IBSP	0.9	0.018	FUT8	0.7	0.039
FIGF	-1.2	0.001	CCND1	1.0	0.018	BRCA1	0.8	0.042
IL19	1.3	0.002	GADD45G	-0.9	0.020	LIFR	-0.7	0.042
IFNA17	-1.4	0.002	IRS1	0.9	0.020	CDH1	0.9	0.042
BAIAP3	1.2	0.002	PRLR	1.0	0.023	ZBTB16	-0.9	0.043
ACVR1C	-1.0	0.002	GNAS	0.9	0.023	RNF43	0.8	0.044
JUN	-1.1	0.002	CACNG1	0.9	0.023	IL13	-0.7	0.044
IGF1	-1.0	0.002	B2M	0.8	0.024	HSPB1	0.8	0.044
INHBA	1.1	0.002	CCNE2	1.0	0.024	OSM	-0.7	0.044
SOCS3	-1.0	0.002	SFRP4	-0.9	0.025	RXRG	-0.8	0.044
MAPT	1.4	0.003	RASGRP1	1.0	0.025	RUNX1T1	-0.7	0.044
MYB	1.3	0.003	WHSC1L1	0.8	0.026	EFNA3	0.8	0.047
PROM1	1.4	0.006	DKK2	-0.8	0.026	PBX3	0.7	0.047
CAMK2B	1.2	0.006	PIK3R3	0.9	0.026	GADD45B	-0.7	0.047
CDC6	1.2	0.007	ETV7	0.9	0.027	SOST	-0.9	0.049
SHC4	1.2	0.008	SSX1	-0.9	0.027	FZD3	0.7	0.049
SPRY1	-0.9	0.008	RAC3	0.8	0.028	IDH2	0.8	0.049
BRIP1	1.1	0.008	HDAC11	0.8	0.028			
COL6A6	-1.0	0.008	WIF1	-1.0	0.028			

Para tener una visión más global de los resultados obtenidos, con la herramienta online *DAVID Gene Funtional Classification Tool*, se ha hecho una clasificación según la función de los genes DE (Tabla 22). De los 112 genes DE, 88 no son clasificados mientras que 22 de esos genes se agrupan en 5 clústeres diferentes. El grupo 1 son miembros de la familia Wnt (WNT7B, Log2 FC 1.1; WNT16, Log2 FC -1.4; WNT2, Log2 FC 0.9; WNT11, Log2 FC -0.8), el grupo 2, receptores de membrana (PRLR, Log2 FC 1; IL1R2, Log2 FC -0.8; LIFR, Log2 FC -0.7; LEPR, Log2 FC -0.9; CSF3R, Log2 FC 0.8), grupo 3, fosfolipasas (PLA2G5, Log2 FC -0.8; PLA2G2A, Log FC -1.3; PLA2G10, Log2 FC 0.9; PLA2G4E, Log2 FC -1.5), grupo 4, subunidades de los canales de calcio voltaje dependiente (CACNG4, Log2 FC 2; CACNA2D3, Log2 FC -0.9; CACNG1, Log2 FC 0.9; CACNB3, Log2 FC 0.8; CACNA1G, Log2 FC -0.7) y el grupo 5, subunidades de integrinas (ITGA2, Log2 FC 0.8; ITGB6, Log2 FC 1.6; ITGB4, Log2 FC 0.8; ITGA7, Log2 FC -1.2).

Tabla 22. Clasificación de los 112 genes DE con la herramienta *DAVID Gene Funtional Classification Tool*.

Clasificación funcional de los 112 genes DE		
Grupo de genes 1		
1	WNT7B	<i>Wnt family member 7B(WNT7B)</i>
2	WNT16	<i>Wnt family member 16(WNT16)</i>
3	WNT2	<i>Wnt family member 2(WNT2)</i>
4	WNT11	<i>Wnt family member 11(WNT11)</i>
Grupo de genes 2		
1	PRLR	<i>prolactin receptor(PRLR)</i>
2	IL1R2	<i>interleukin 1 receptor type 2(IL1R2)</i>
3	LIFR	<i>leukemia inhibitory factor receptor alpha(LIFR)</i>
4	LEPR	<i>leptin receptor(LEPR)</i>
5	CSF3R	<i>colony stimulating factor 3 receptor(CSF3R)</i>
Grupo de genes 3		
1	PLA2G5	<i>phospholipase A2 group V(PLA2G5)</i>
2	PLA2G2A	<i>phospholipase A2 group IIA(PLA2G2A)</i>
3	PLA2G10	<i>phospholipase A2 group X(PLA2G10)</i>
4	PLA2G4E	<i>phospholipase A2 group IVE(PLA2G4E)</i>
Grupo de genes 4		
1	CACNG4	<i>calcium voltage-gated channel auxiliary subunit gamma 4(CACNG4)</i>
2	CACNA2D3	<i>calcium voltage-gated channel auxiliary subunit alpha2delta 3(CACNA2D3)</i>
3	CACNG1	<i>calcium voltage-gated channel auxiliary subunit gamma 1(CACNG1)</i>
4	CACNB3	<i>calcium voltage-gated channel auxiliary subunit beta 3(CACNB3)</i>
5	CACNA1G	<i>calcium voltage-gated channel subunit alpha1 G(CACNA1G)</i>
Grupo de genes 5		
1	ITGA2	<i>integrin subunit alpha 2(ITGA2)</i>
2	ITGB6	<i>integrin subunit beta 6(ITGB6)</i>
3	ITGB4	<i>integrin subunit beta 4(ITGB4)</i>
Sin grupo		
88 genes del estudio no pueden ser agrupados		

5.3.2.2. Enfermedad residual (RD_T2) versus metástasis (Mtx_T3).

Las diferencias en cuanto al tumor residual frente al primer nicho metastásico presentan una activación de los genes DE en la evolución del tumor, ya que todos los genes se encuentran sobreexpresados en el tumor metastásico. Los genes DE son 30 y quedan reflejado en la tabla 23 y en la figura 32, junto con su Log2 FC y el p-valor ajustado. Al igual que en el apartado anterior, se va a describir la función de los 10 primeros genes con el p-valor más bajo y su relación con el proceso que queda reflejado según nuestros resultados.

El gen DE con más poder estadístico es **MAPK8IP2** que se encuentra menos expresado en la RD (Log2 FC -2.8, $p=3.98E^{-05}$). El producto de este gen está envuelto en la regulación de la vía de señalización de la quinasa amino terminal c-Jun (RefSeq, 2017). La variante de un ayustamiento en el exón 2 de dicho gen se ha relacionado con el incremento de propiedades de células stem/progenitoras aumentando procesos de desdiferenciación a través de la activación de la vía JNK (125). Que haya un aumento de procesos de desdiferenciación por el incremento de propiedades de células progenitoras es una de las características de células con capacidad metastásica, por lo tanto, los resultados son coherentes.

El gen **IL22RA2**, al igual que el resto de los genes DE, se encuentra expresado a la baja en la RD con respecto a la metástasis (Log2 FC -2.71, $p=1.14E^{-04}$). Es la subunidad Alpha 2 de la interleuquina 22 (IL22), está relacionado con regulación de la respuesta inflamatoria y está implicado en la regulación de la tumorigenesis en colon (RefSeq, 2013). Este, juntos a otros genes, han sido identificados como re

EL gen **LAMB4**, expresado a la baja en el tumor residual (Log2 FC -2.5, $p=0.0014$), es la subunidad Beta 4 de la Laminina. Se piensa que la laminina media en la organización y migración de células en los tejidos durante el desarrollo embrionario por la interacción de otros componentes de la matriz celular. La sobreexpresión de este gen ha sido relacionada con disminución de la supervivencia en tumores de mama ER+, y el efecto contrario en los tumores ER- (33). El aumento de la expresión de este gen en el tumor

metastásico podría deberse al proceso metastásico en sí donde es necesaria una reorganización de las células en el nuevo tejido.

El gen *CAMK2B* expresado a la baja en la RD (Log2 FC -2.5, p=0.0014), codifica para una proteína quinasa dependiente de Ca²⁺/Calmodulina involucrada en la señalización por calcio, muy importante en la plasticidad de la sinopsis glutamatérgica (RefSeq, 2014). En líneas celulares de cáncer de mama, se ha demostrado la hipermetilación de dicho gen y consecuente bajada en la expresión génica (35). En un estudio previo donde se identificaban los genes DE entre pacientes que obtenían y no obtenían pCR tras el tratamiento con NAC, se vio que este gen se encontraba más expresado en aquellos tumores que respondían al tratamiento con NAC (126). En base a nuestros resultados, el gen *CAMK2B*, que también se encuentra más expresado en el tumor primario antes del tratamiento que en la RD (Log2 FC 2, p=0.006), podría seleccionarse negativamente tras el tratamiento con NAC.

EL gen *PPP2R2C*, que se encuentra menos expresado en la RD con respecto a la metástasis (Log2 FC -2.4, p=0.0026), codifica para la subunidad reguladora de la fosfatasa 2, y dicha fosfatasa es una de las 4 Serina/Treonina fosfatasas más importantes. Está implicado en el control negativo del crecimiento celular y división (RefSeq, 2008). En cáncer de próstata se ha visto que una pérdida en la expresión de este gen se asocia a un aumento de sufrir metástasis (127). Debido a que en la RD de nuestra cohorte se encuentra expresada a la baja y son pacientes que han sufrido metástasis, podría ser un marcador de procesos metastásicos en RD de tumores de mama.

EL gen *CACNA1D*, menos expresado en la RD (Log2 FC -2.5, p=0.007), codifica para la subunidad alfa 1D del canal regulado por voltaje de calcio. Está involucrada en muchos procesos de los cuales se requiere de la presencia de calcio, como la contracción del musculo, lo relacionado con hormonas y neurotransmisores y expresión génica (RefSeq, 2012). Entre las funciones de los canales de calcio regulado por voltaje en el cáncer está la de mitogenesis, proliferación, diferenciación, apoptosis y metástasis (128). En cáncer de hígado y pulmón se ha descrito una bajada en la expresión de *CACNA1D* en el tejido tumoral en comparación con el tejido normal, lo que ellos determinaron como un gen

biomarcador de supresión de tumores en estos tipos de cáncer (128). En base a nuestros resultados, podemos decir que la alta expresión de este gen se asocia a tumores metastásicos en cáncer de mama que han sido tratados con NAC.

El gen *ETV4*, expresado a la baja en la RD (Log FC -2, $p=0.0082$), es un factor de transcripción que se expresa en las células epiteliales de la glándulas mamarias, que está relacionado con la super vía de la desregulación transcripcional en cáncer (129). Está relacionado con procesos de proliferación, migración e invasión (130). La sobreexpresión de este gen se ha visto asociada a un alto riesgo de metástasis a distancia en tumores de mama TN en cultivos celulares (131) y en muestras de cáncer de mama de todos los subtipos (130). En nuestros resultados, se trataría de un marcador de metástasis tras el tratamiento con NAC.

EL gen *RET*, que se encuentra expresado a la baja en la RD (Log₂ FC -2.3, $p=0.0082$), es un protooncogén estrechamente relacionado con roles de proliferación celular, crecimiento, migración y supervivencia (RefSeq, 2017). Todas estas características se asocian con funciones necesarias para el proceso metastásico, su alta expresión en el tumor metastásico de nuestra cohorte tiene coherencia.

El gen *WNT16*, expresado a la baja en la RD (Log₂ FC -1.8, $p=0.012$), está implicado en oncogénesis y en diferentes procesos del desarrollo (RefSeq, 2008). En cáncer de mama, la subida en la expresión de este gen se ve inducido al tratamiento con taxanos, se ha asociado a progresión y quimiorresistencia (34). En nuestros resultados, este gen, que se encuentra más activado en la RD residual cuando es comparado con el tumor primario (Log₂ FC -1.4, $p=7E^{-04}$) también se ve más expresados progresivamente en la RD tras el tratamiento con NAC y expresado, más aún, en la metástasis con respecto a la RD. Podríamos hablar de un marcador con aumento de expresión que evoluciona desde el tratamiento con NAC a fenotipos más agresivos como la enfermedad metastásica.

El gen *NPM2*, expresado a la baja en la RD con respecto al tumor metastásico (Log₂ FC -3.1, $p=0.012$), es miembro de la familia de las chaperonas nucleolares. Esta familia de proteínas está implicada en diversas funciones celulares como el procesamiento del RNA ribosomal, el ciclo celular, la duplicación del centrosoma y el control de procesos para mantener la estabilidad genómica (132) (133).

Tabla 23. Genes DE entre RD_T2 y Mtx_T3

Gen	Log2 FC	p-valor ajustado
MAPK8IP2	-2.8	3.98E-05
IL22RA2	-2.7	1.145E-04
LAMB4	-2.5	0.0014
CAMK2B	-2.5	0.0014
PPP2R2C	-2.4	0.0026
CACNA1D	-1.9	0.0070
ETV4	-2.0	0.0082
RET	-2.3	0.0082
WNT16	-1.8	0.0127
NPM2	-3.1	0.0127
MLLT4	-1.6	0.0127
IBSP	-1.8	0.0151
SFN	-1.9	0.0151
CCNE1	-2.2	0.0321
PLA2G3	-1.6	0.0321
CDC25A	-1.8	0.0321
ITGB6	-2.0	0.0321
RASGRP1	-2.0	0.0321
HIST1H3H	-1.7	0.0375
ITGA3	-1.4	0.0395
GRIN2A	-1.4	0.0395
LAMA5	-1.2	0.0395
FGF1	-1.6	0.0422
COL2A1	-1.7	0.0425
MMP9	-1.8	0.0495
RAC3	-1.5	0.0495
CREB3L4	-1.5	0.0495
HHIP	-1.3	0.0495
WT1	-1.6	0.0495
IL12RB2	-1.5	0.0495

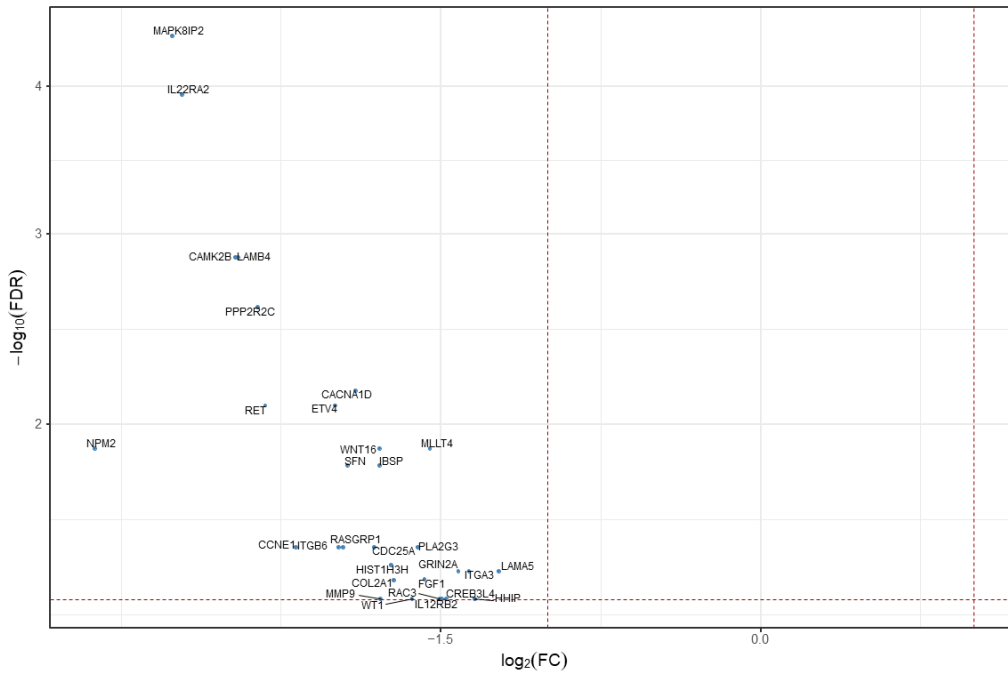


Figura 32. En el *volcano plot* se representan todos los genes según el p-valor ajustado obtenido en el AED (eje de ordenadas) y el valor del Log₂ FC (eje de abscisas). Con las líneas punteadas se representan diferentes niveles de p-valor y la zona del *volcano* donde se posicionará el gen corresponderá con un p-valor concreto y un valor de FC determinado. En este caso, los genes que queden a la derecha del punto 0 del eje de abscisas serán los genes sobreexpresados en el RD_T2 y los que queden en el lado izquierdo serán los sobreexpresados en el Mtx_T3.

5.3.2.3. Tumor primario (Pre_T1) versus metástasis (Mtx_T3)

Los genes DE entre el tumor primario y la metástasis son *FOS*, *WNT16*, *GRIN1*, *CAMK2B*, *IFNA17*, *NR4A1*, *IL1A*, *NTRK2*, *NPM2*, *IFNA7* y *LAMB4*. Todos ellos se encuentran expresados a la baja en el tumor primario al compararlo con la metástasis (Tabla 24 y Figura 33).

La función o implicación en cáncer de muchos de estos genes (*FOS*, *WNT16*, *CAMK2B*, *NR4A1*, *NPM2* y *LAMB4*) ya han sido explicadas en los apartados anteriores por lo que la descripción en este apartado será más breve.

En primer lugar, el gen *FOS*, que se encuentra menos expresado en el tumor primario frente a la metástasis (Log₂ FC -3.5, p=6.28E⁻⁰⁶), como hemos comentado con anterioridad, es un protooncogén que está involucrado en procesos de resistencia a la apoptosis y potencial marcador de resistencia a terapia (28). También se encuentra expresado al alza en la enfermedad residual con respecto al tumor primario (Log FC - 3.55, p=2.57E⁻²³), por lo que su función como biomarcador de resistencia a terapia con NAC parece clara.

El gen *WNT16*, expresado a la baja en el tumor primario (Log2 FC -2.4, p=5.06E⁻⁰⁴), se relaciona con en oncogénesis, progresión y quimioresistencia a taxanos (34). Que dicho gen se encuentre más expresado en la metástasis tiene relación directa con la progresión y resistencia a los taxanos, que es parte del tratamiento con NAC.

El gen *GRIN1*, expresado a la baja en el tumor primario (Log2 FC -2.1, p=0.001), es miembro de los canales de receptores de glutamato. Juega un papel importante en la plasticidad de la sinapsis (RefSeq, 2008). En nuestra cohorte se ve relacionada con fenotipo metastásico con respecto al tumor primario.

EL gen *CAMK2B*, se expresa a la baja en el tumor primario antes del tratamiento (Log2 FC -2.4, p=0.001). Este gen ha sido previamente relacionado con respuesta al tratamiento con NAC (126). En la comparativa entre tumor primario y metástasis, la expresión de este gen se ve aumentado en la metástasis en nuestra cohorte.

EL gen *IFNA17*, menos expresado en el tumor primario (Log2 FC -2.1, p=0.002), codifica para el interferón alfa 17, implicado en procesos activación de citoquinas. Al igual que el resto de estos genes, podría tratarse de un marcador de metástasis al tratamiento con NAC.

EL gen *NR4A1*, expresado a la baja en el tumor primario frente a la metástasis (Log2 FC -2.5, p=0.002), es un factor de transcripción nuclear, relacionado con procesos de progresión tumoral, EMT y migración celular (30). Que se encuentre sobreexpresado en la metástasis con respecto al tumor primario en nuestra cohorte es coherente con lo descrito en la literatura.

El gen *IL1A* es, junto con *NTRK2*, de los genes DE entre el tumor primario y la metástasis, que no los están en las demás comparativas. Ambos genes se encuentran expresados a la baja en el tumor primario con respecto a la metástasis (Log2 FC -2.7, p=0.004 y Log2 FC -2, p=0.004, respectivamente). El *IL1A* codifica para la interleuquina 1. Esta citoquina está envuelta en respuesta inmune, procesos inflamatorios y hematopoyesis (RefSeq, 2008). El gen *NTRK2* codifica para un miembro de la familia de tirosina quinasa TrkA. La señalización a través de esta quinasa está relacionada con diferenciación celular (RefSeq, 2014).

EL gen *NPM2*, se encuentra más expresado en las muestras de tejido metastásico que en los tumores primarios antes de ser tratados (Log2 FC -3, p=0.02). Como hemos comentado con anterioridad, forma parte de la familia de las chaperonas y está involucrado en proceso de ciclo celular y estabilización de estabilidad genómica (133).

El gen *IFNA7*, menos expresado, como el resto de los genes, en el tumor primario frente al tumor metastásico (Log2 FC -1.8, p=0.02), codifica para el interferón alfa 7. En cáncer, se encuentra relacionado con vías como la de JAK-STAT y podría estar involucrado en la resistencia mutagénica programada al bloqueo del control inmune (134).

Y, por último, el gen *LAMB4*, expresado a la baja en el tumor primario frente al tumor metastásico (Log2 FC -2, p=0.02), está relacionado con la organización y migración de las células durante el desarrollo embrionario por interacción de otros componentes en la matriz celular. Como comentamos en el apartado anterior donde este gen también se encontraba sobre sobreexpresado en el tumor metastásico, puede deberse al proceso metastásico en sí y a todos los ajustes tisulares que conlleva.

Tabla 24. Genes DE entre Pre_T1 y Mtx_T3

Gen	Log2 FC	p-valor ajustado
FOS	-3.5	6.28E-06
WNT16	-2.4	5.06E-04
GRIN1	-2.1	0.0011
CAMK2B	-2.4	0.0011
IFNA17	-2.1	0.0025
NR4A1	-2.5	0.0025
IL1A	-2.7	0.0046
NTRK2	-2.0	0.0046
NPM2	-3.0	0.0161
IFNA7	-1.8	0.0206
LAMB4	-2.0	0.0276

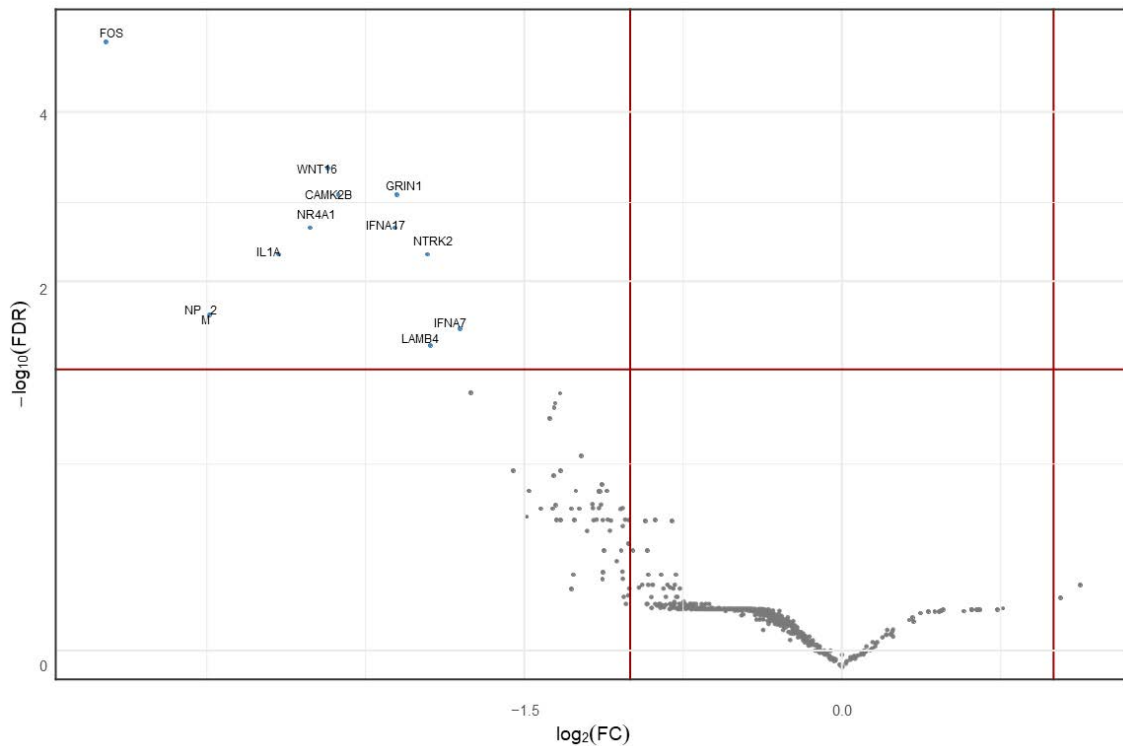


Figura 33. En el *volcano plot* se representan todos los genes según el p-valor ajustado obtenido en el AED (eje de ordenadas) y el valor del Log2 FC (eje de abscisas). Con las líneas punteadas se representan diferentes niveles de p-valor y la zona del *volcano* donde se posiciona el gen corresponderá con un p-valor concreto y un valor de FC determinado. En este caso, los genes que queden a la derecha del punto 0 del eje de abscisas serán los genes sobreexpresados en el Pre_T1 y los que queden en el lado izquierdo serán los sobreexpresados en el Mtx_T3.

5.3.2.4. Tumor primario, enfermedad residual y metástasis

En cuanto a las similitudes entre los diferentes ADE que hemos realizado, Pre_T1/RD_T2/Mtx_T3, existen algunos genes que son compartidos y que tienen un comportamiento diferente entre cada una de las comparativas. En la figura 34 se han especificado aquellos genes que comparten los análisis realizados y cuál es la dirección del cambio en la expresión génica según su valor positivo o negativo del Log2 FC, la referencia siempre es el tiempo más temprano.

Cuando se compara el tumor primario frente a la RD y al tumor metastásico (Figura 33, B), los genes *IFNA17*, *GRIN1*, *FOS*, *NR4A1*, *WNT16* y *CAMK2B* están DE en ambos análisis. El único gen sobreexpresado en el tumor primario con respecto a la RD y a la metástasis es el gen *CAMK2B*. Este gen, podría estar siendo seleccionado negativamente por la NAC, y eliminarse en las clonas celulares sensibles al tratamiento. El resto de los genes (*IFNA17*, *CRIN*, *FOS*, *NR4A1* y *WNT16*), sin embargo, parecen estar siendo marcadores de clonas resistentes a las NAC que podrían haber estado enmascaradas en el tumor

primario y no haberse capturado debido a la heterogeneidad tumoral en la toma de la muestra o ser un cambio introducido por el tratamiento con NAC. Cabe destacar al gen *FOS* (es el gen con más cambio en su expresión de estas comparativas) como biomarcador de resistencia al tratamiento tanto en la RD como en la enfermedad metastásica (Log2 FC -3.6 y -3.5, respectivamente, teniendo como referencia el tumor primario, Pre_T1).

Cuando comparamos la RD con el tumor primario y la enfermedad metastásica, los genes DE compartidos son *RAC3*, *FGF1*, *IBSP*, *WNT16*, *CAMK2B*, *ITGB6*, *MAPK8IP2*, *RASGRP1* y *MLLT4*. En este caso, todos los genes están más expresado en el tumor primario y en la enfermedad metastásica, mostrando una semejanza en los perfiles de expresión génica de estos genes entre el tumor primario y la metástasis, a excepción de *WNT16*, que presenta una expresión que va evolucionando positivamente tras el tratamiento con NAC siendo mayor en la RD con respecto al tumor primario, y en la enfermedad metastásica con respecto a la RD (Figura 34, C).

En el caso de la metástasis frente al tumor primario y a la RD, los genes que se comparten son *LAMB4*, *WNT16*, *NPM2* y *CAMK2B*. Estos cuatro genes se encuentran más expresados en la enfermedad metastásica con respecto al tumor primario y a la RD, seleccionándose positivamente en el proceso metastásico tras el tratamiento con NAC. El gen con mayor cambio de expresión tanto cuando se compara con el tumor primario como con la RD es *NPM2*. (Figura 34, D)

Cuando analizamos todas los datos a la vez, vemos que los genes que comparten todas las comparativas son *WNT16* y *CAMK2B*. En el caso del gen *WNT16*, ya hemos comentado esa posible selección positiva de su expresión tras el tratamiento con NAC, debido al aumento de expresión génica progresiva entre Pre_T1, RD_T2 y Mtx_T3. En el caso de *CAMK2B* vemos una caída de la expresión génica tras el tratamiento, tanto en la RD como en el tumor metastásico, proponiendo un nuevo marcador de sensibilidad a la NAC (Figura 34, E).

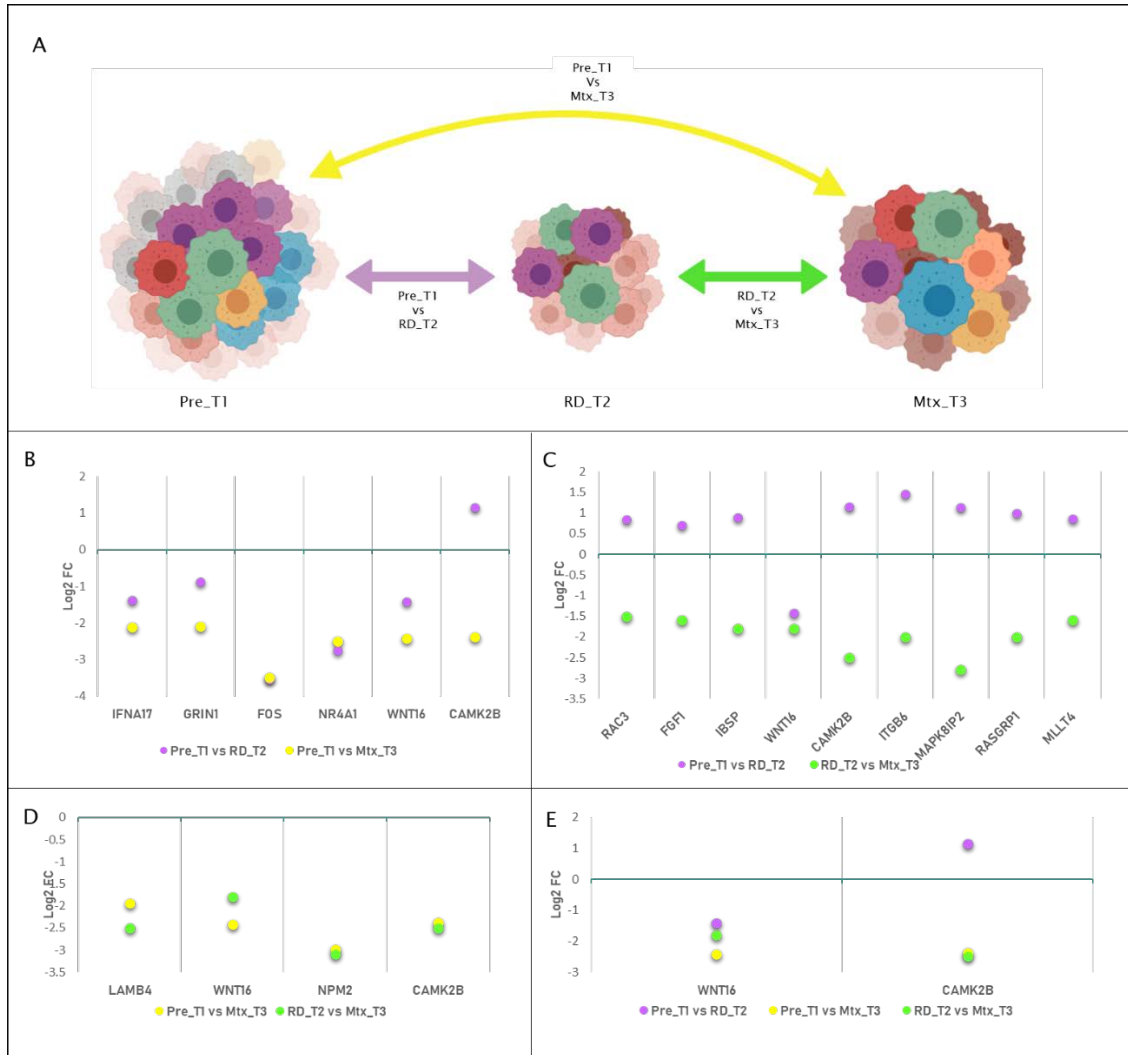


Figura 34. A) Representación gráfica de las comparaciones realizadas en base a los resultados obtenidos en los diferentes AED. B) Genes DE comunes a la comparación de Pre_T1vsRD_T2 y Pre_T1vsMtx_T3 distribuidos según el Log₂ FC. C) Genes DE comunes a la comparación de Pre_T1vsRD_T2 y RD_T2vsMtx_T3 distribuidos según el Log₂ FC. D) Genes DE comunes a la comparación de Pre_T1vsMtx_T3 y RD_T2vsMtx_T3 distribuidos según el Log₂ FC. E) Genes DE comunes a la comparación de Pre_T1vsRD_T2, Pre_T1vsMtx_T3 y RD_T2vsMtx_T3 distribuidos según el Log₂ FC.

5.4. Discusión

El principal objetivo de este capítulo es definir posibles marcadores de sensibilidad y resistencia al tratamiento con NAC, así como, determinar las similitudes y diferencias entre el tumor residual tras la NAC y el primer nicho metastásico. En base a los resultados obtenidos, hemos definido una serie de genes que se expresan diferencialmente y que varía desde el tumor primario hasta la enfermedad metastásica, dando una idea de los cambios seleccionados con la NAC.

Entre el tumor primario y la enfermedad residual hay una serie de genes DE que codifican productos con funciones de diversa índole, factores de transcripción, como es el caso de FOS y JUN que forman el complejo del factor de transcripción de AP1 relacionado con la regulación de la proliferación y de la apoptosis, receptores de hormonas como NR4A1 y NR4A3, y otros receptores de membrana como PRLR, IL1R2, LIFR, LEPR y CSF3R, varios miembros de la familia Wnt (WNT7B, WNT16, WNT2 y WNT11), genes relacionados con los canales de calcio voltaje-dependiente (CACNG4, CACNA2D3, CACNG1, CACNB3 y CACNA1G) y varios tipos de integrinas (ITGA2, ITGB6, ITGB4 y ITGA7), entre otros. Todos aquellos cambios que se seleccionan positivamente tras el tratamiento con NAC proporcionan resistencia y caracterizan a una población con los mecanismos moleculares necesarios para hacer frente al tratamiento citotóxico. De los genes DE entre el tumor primario y la RD identificados en este estudio, *PLAU*, *STAT1* y *WNT11* ya habían sido descritos como genes con cambios en su expresión tras el tratamiento neoadyuvante en estudios anteriores (1) (32). El perfil de expresión de dichos genes se corresponde con el perfil que hemos obtenido en nuestros análisis, quedando respaldado su papel en la sensibilidad al tratamiento con NAC (en el caso de *PLAU* y *STAT1*) o en la resistencia a este (en el caso de *WNT11*).

En algunos casos, los cambios encontrados en nuestra cohorte antes y después del tratamiento se mantienen hasta el primer nicho metastásico. Un ejemplo de ello es el gen *FOS*, en estudios previos se ha demostrado que la disminución de su expresión génica está relacionado con fenotipos sensibles a agentes quimioterapéuticos (28). En nuestra cohorte el comportamiento de la expresión del gen *FOS* es el mismo que en lo descrito en la literatura, pero, además, es el gen que más cambia su expresión génica

entre dicho tumor primario y el primer nicho metastásico. Sin embargo, no hay diferencias significativas entre la RD y la metástasis, lo que podría significar que una vez que el tumor es sometido al tratamiento con NAC se seleccionan aquellas clonas con alta expresión de *FOS* (marcado por su función de regulación de la apoptosis y supervivencia) y esta selección positiva le confiere ventajas evolutivas que se mantiene hasta el primer nicho metastásico. El mismo perfil de expresión que *FOS* siguen los genes *NR4A1*, *GRIN1* y *IFNA17*, menos expresado en el tumor primario con respecto a la RD y menos expresado entre el tumor primario con respecto a la metástasis, sin diferencias significativas entre la RD y la metástasis. Por lo tanto, proponemos a los genes *FOS*, *NR4A1*, *GRIN1* y *IFNA17* como marcadores de resistencia tras el tratamiento con NAC que, además, podría definir a una población celular con capacidad metastásica. En la enfermedad residual se observa una bajada en la expresión de todos los genes que se encuentran DE cuando se compara con la metástasis. Podríamos suponer que hay una serie de genes, *RAC3*, *FGF1*, *IBSP*, *ITGB6*, *MAPK8IP2*, *RASGRP1* y *MLLT4*, que, en el tumor metastásico o previamente a la invasión del órgano diana, se están sobreexpresando y que pueden estar relacionados con la agresividad propia del tumor desde el inicio del tratamiento con NAC, ya que estos genes también se sobreexpresan más en el tumor primario y en la enfermedad metastásica que en la RD, sin diferencias significativas entre el tumor primario y el tumor metastásico.

En cuanto a la enfermedad metastásica, podríamos definir que los genes *LAMB4* y *NPM2* son biomarcadores de recaída a distancia cuando son comparados con el tumor primario y con la RD. Dichos genes se sobreexpresan en el tumor metastásico, pero no en el tumor primario ni en la RD. Estudios donde se profile la expresión génica de ambos genes en el tejido de la metástasis ayudarían a definir si su expresión es específica del órgano diana por una posible infiltración del tejido adyacente, o si por el contrario es un marcador exclusivo de metástasis en el cáncer de mama.

El gen *WNT16*, que está asociado a progresión y quimiorresistencia inducida por taxanos (34), tiene un perfil de expresión donde tras el tratamiento con NAC es seleccionado positivamente y esta selección es progresiva aumentando la expresión génica desde el tumor primario a la RD y de la RD al tumor metastásico, quizás seleccionando super

clonas tumorales. Por otro lado el gen *CAMK2B*, involucrado en la señalización con calcio, en estudios anteriores ha sido asociado a respuesta al tratamiento con NAC (35). En nuestra cohorte, este gen se encuentra más expresado en el tumor primario con respecto a la RD tras la NAC, respaldando los resultados previos. Según nuestro análisis longitudinal, la expresión génica de la *CAMK2B* aumenta significativamente mucho más en la metástasis que incluso en el tumor primario, donde tiene mayor expresión que en la RD.

Hay que mencionar, también, las limitaciones que tiene el estudio. Las cohortes donde se han establecido las diferencias en la expresión génica de los 730 genes analizados no tienen el mismo número de pacientes, 70 para la comparativa Pre_T1vsRD_T2 y 22 para la comparativa Pre_T1vsMtx_T3 y RD_T2vsMtx_T3, todas ellas pareadas. En este caso, en la comparativa Pre_T1vsRD_T2 quizás se han capturado mejor (en términos estadísticos) los cambios encontrados en los genes analizados.

Otra cuestión para tener en cuenta son los órganos donde ha tenido lugar el primer nicho metastásico tras la RD. No hemos seleccionado un órgano concreto diana de metástasis, sino que hemos capturado marcadores globales y no específicos de tejido. Además, entre las muestras de Mtx_T3, hay muestras de piel pertenecientes al tejido cercano a la piel de la mama, lo que podría considerarse recaída local y no a distancia, aunque son el menor de los casos.

5.5. CONCLUSIONES

En base a nuestros resultados y los objetivos propuestos, se han identificado alteraciones en la expresión génica tanto comunes como específicos en el tumor primario al compararlo con la enfermedad residual y el primer nicho metastásico. Se muestra una inactivación de la expresión génica de los genes *IFNA17*, *GRIN1*, *FOS* *NRA41* y *WNT16* en el tumor primario frente a la enfermedad residual y a la enfermedad metastásica, donde se encuentran más expresados. La sobreexpresión de dichos genes en la RD podría proporcionar una serie de nuevos marcadores en los que seguir indagando para proponer nuevas dianas terapéuticas en la enfermedad residual una vez que el tratamiento con NAC no ha sido efectivo. Con respecto a la metástasis, nuestros resultados proponen 4 genes DE entre la metástasis versus la enfermedad residual y el

tumor primario. *LAMB4* y *NPM2* como genes DE con mayor expresión únicamente en la metástasis, *CAMK2B* como un gen que se selecciona negativamente tras la NAC en la enfermedad residual pero que en el tumor metastásico recupera su actividad, y *WNT16* como un posible marcador de evolución desfavorable del tumor tras la NAC.

6. Conclusiones

Si sabes que estás en el camino correcto, si lo sientes,
entonces... no importa lo que digan.

Barbara McClintock

Los resultados de este trabajo de investigación han permitido la caracterización global de los cambios a nivel de expresión génica de una serie de genes relacionados estrechamente con el proceso tumoral tras el tratamiento con NAC. Además, hemos obtenido unos resultados muy específicos en cuanto a la predicción a la respuesta al tratamiento al inicio de la enfermedad. Las conclusiones principales de esta Tesis Doctoral son las siguientes:

1. Se ha identificado una firma predictiva basada en las diferencias biológicas encontradas entre las pacientes que obtienen pCR y las que no, independiente del subtipo, y que mejora a las actuales herramientas de predicción. Esta mejora en la identificación de las pacientes respondedoras versus no respondedoras es más notable en las pacientes ER+; esto es especialmente relevante porque estas son las pacientes que menos responden a la NAC y por tanto las que más se beneficiarían de una discriminación más específica.
2. Entre el tumor primario de las pacientes no respondedoras y la enfermedad residual, hay un espectro de diferencias biológicas a nivel de expresión génica que abarcan la función de muchos tipos de genes. Destacan el papel de la expresión del proto-oncogén FOS y de los receptores de membrana NR4A1 y NR4A3, que son importantes en la regulación de la proliferación y procesos apoptóticos.
3. La relación que encontramos entre el tumor primario antes de ser tratado, la enfermedad residual después del tratamiento y los primeros nichos metastásicos, muestran una serie de alteraciones de las que se podría derivar la caracterización de la evolución tumoral tras la NAC. Se muestra una inactivación de la expresión génica de los genes *IFNA17*, *GRIN1*, *FOS*, *NRA41* y *WNT16* en el tumor primario frente a la enfermedad residual y a la enfermedad metastásica, donde se encuentran más expresados. La sobreexpresión de dichos genes en la RD podría proporcionar una serie de nuevos marcadores que seguir investigando para proponer nuevas dianas terapéuticas en la enfermedad residual una vez que el tratamiento con NAC no ha sido efectivo. Con respecto a la metástasis, nuestros resultados proponen cuatro genes DE entre la metástasis versus la

enfermedad residual y el tumor primario: *LAMB4* y *NPM2* como genes DE con mayor expresión únicamente en la metástasis, *CAMK2B* como un gen que se selecciona negativamente tras la NAC en la enfermedad residual pero que en el tumor metastásico recupera su actividad, y *WNT16* como un posible marcador de evolución desfavorable del tumor tras la NAC.

Los resultados de la presente tesis doctoral representan un avance en el conocimiento molecular de los distintos momentos del cáncer de mama tratado con NAC desde el tumor primario a la recaída a distancia. Los modelos predictivos de pCR, así como las moléculas que definen diferencialmente a la RD y a la metástasis son la base para la implementación de la medicina de precisión y el descubrimiento de nuevas dianas terapéuticas a la resistencia en esta enfermedad de alta prevalencia, y cuya recaída implica el 90% de las muertes relacionadas con ella.

7. Conclusions

The more we know, the more we realize there is to know.

Jennifer Doudna

The results of this research work have allowed the global characterization of the changes at the level of gene expression of a series of genes closely related to the tumor process after treatment with NAC. In addition, we have obtained very specific results in terms of predicting the response to treatment at the onset of the disease. The main conclusions of this Doctoral Thesis are the following:

1. A predictive signature has been identified based on the biological differences found between patients who obtain pCR and those who do not, independent of the subtype, and which improves the current prediction tools. This improvement in the identification of responding versus non-responding patients is more notable in ER+ patients; this is especially relevant because these are the patients who least respond to NAC and therefore those who would benefit the most from a more specific discrimination.
2. There is a spectrum of biological differences at the level of gene expression between the primary tumor and the residual disease that encompass the function of many types of genes. Among them, the role of the expression of the proto-oncogene *FOS* and the membrane receptors *NR4A1* and *NR4A3*, which are important in the regulation of proliferation and apoptotic processes, seems particularly important for defining the resistance to NAC.
3. The relationship that we have delineated between the primary tumor, the residual disease after the treatment and the first metastatic niche, shows a series of alterations that characterize the tumor evolution after NAC. An inactivation of the gene expression of the *IFNA17*, *GRIN1*, *FOS*, *NRA41*, and *WNT16* genes is shown in the primary tumor compared with residual disease and metastatic disease, where they are more expressed. The overexpression of these genes in RD could provide a panel of new markers worth to be further investigated in order to propose new therapeutic targets in residual disease once treatment with NAC has not been effective. Regarding the metastasis stage, our results propose four DE genes between metastasis, residual disease and primary tumor: *LAMB4* and *NPM2* as DE genes with higher expression only in metastasis, *CAMK2B* as a gene that is negatively selected after NAC in residual disease but which recovers

its activity in the metastatic tumor, and *WNT16* as a possible marker of unfavorable evolution of the tumor after NAC.

The results of this Doctoral Thesis represent an advance in the molecular knowledge of the different moments of breast cancer treated with NAC from the primary tumor to the distant relapse. The predictive models of pCR, as well as the molecules that differentially define RD and metastasis are the basis for the implementation of precision medicine and the discovery of new therapeutic targets for resistance in this highly prevalent disease, whose relapse involves 90% of deaths related to it.

8. Referencias

1. Chica-Parrado MR, Godoy-Ortiz A, Jiménez B, Ribelles N, Barragan I, Alba E. Resistance to Neoadjuvant Treatment in Breast Cancer: Clinicopathological and Molecular Predictors. *Cancers*. 22 de julio de 2020;12(8):2012.
2. Prat A, Fan C, Fernández A, Hoadley KA, Martinello R, Vidal M, et al. Response and survival of breast cancer intrinsic subtypes following multi-agent neoadjuvant chemotherapy. *BMC Med*. diciembre de 2015;13(1):303.
3. von Minckwitz G, Untch M, Blohmer J-U, Costa SD, Eidtmann H, Fasching PA, et al. Definition and Impact of Pathologic Complete Response on Prognosis After Neoadjuvant Chemotherapy in Various Intrinsic Breast Cancer Subtypes. *JCO*. 20 de mayo de 2012;30(15):1796–804.
4. Gentile LF, Plitas G, Zabor EC, Stempel M, Morrow M, Barrio AV. Tumor Biology Predicts Pathologic Complete Response to Neoadjuvant Chemotherapy in Patients Presenting with Locally Advanced Breast Cancer. *Annals of Surgical Oncology*. diciembre de 2017;24(13):3896–902.
5. Ohara AM, Naoi Y, Shimazu K, Kagara N, Shimoda M, Tanei T, et al. PAM50 for prediction of response to neoadjuvant chemotherapy for ER-positive breast cancer. *Breast Cancer Res Treat*. febrero de 2019;173(3):533–43.
6. Lips EH, Mulder L, de Ronde JJ, Mandjes IAM, Koolen BB, Wessels LFA, et al. Breast cancer subtyping by immunohistochemistry and histological grade outperforms breast cancer intrinsic subtypes in predicting neoadjuvant chemotherapy response. *Breast Cancer Res Treat*. julio de 2013;140(1):63–71.
7. Díaz-Redondo T, Lavado-Valenzuela R, Jimenez B, Pascual T, Gálvez F, Falcón A, et al. Different Pathological Complete Response Rates According to PAM50 Subtype in HER2+ Breast Cancer Patients Treated With Neoadjuvant Pertuzumab/Trastuzumab vs. Trastuzumab Plus Standard Chemotherapy: An Analysis of Real-World Data. *Front Oncol*. 5 de noviembre de 2019;9:1178.

8. Alba E, Lluch A, Ribelles N, Anton-Torres A, Sanchez-Rovira P, Albanell J, et al. High Proliferation Predicts Pathological Complete Response to Neoadjuvant Chemotherapy in Early Breast Cancer. *The Oncologist*. 2 de enero de 2016;21(2):150–5.
9. Zhang G, Xie W, Liu Z, Lin C, Piao Y, Xu L, et al. Prognostic Function of Ki-67 for Pathological Complete Response Rate of Neoadjuvant Chemotherapy in Triple-negative Breast Cancer. *Tumori Journal*. marzo de 2014;100(2):136–42.
10. Cabrera-Galeana P, Muñoz-Montaña W, Lara-Medina F, Alvarado-Miranda A, Pérez-Sánchez V, Villarreal-Garza C, et al. Ki67 Changes Identify Worse Outcomes in Residual Breast Cancer Tumors After Neoadjuvant Chemotherapy. *The Oncologist*. junio de 2018;23(6):670–8.
11. Gluz O, Kolberg-Liedtke C, Prat A, Christgen M, Gebauer D, Kates R, et al. Efficacy of deescalated chemotherapy according to PAM50 subtypes, immune and proliferation genes in triple-negative early breast cancer: Primary translational analysis of the WSG-ADAPT-TN trial. *Int J Cancer*. enero de 2020;146(1):262–71.
12. Cortazar P, Geyer CE. Pathological Complete Response in Neoadjuvant Treatment of Breast Cancer. *Ann Surg Oncol*. mayo de 2015;22(5):1441–6.
13. Spring L, Greenup R, Niemierko A, Schapira L, Haddad S, Jimenez R, et al. Pathologic Complete Response After Neoadjuvant Chemotherapy and Long-Term Outcomes Among Young Women With Breast Cancer. *Journal of the National Comprehensive Cancer Network*. 1 de octubre de 2017;15(10):1216–23.
14. Huober J, Holmes E, Baselga J, de Azambuja E, Untch M, Fumagalli D, et al. Survival outcomes of the NeoALTTO study (BIG 1-06): updated results of a randomised multicenter phase III neoadjuvant clinical trial in patients with HER2-positive primary breast cancer. *European Journal of Cancer*. septiembre de 2019;118:169–77.

15. Prat A, Bianchini G, Thomas M, Belousov A, Cheang MCU, Koehler A, et al. Research-Based PAM50 Subtype Predictor Identifies Higher Responses and Improved Survival Outcomes in HER2-Positive Breast Cancer in the NOAH Study. *Clinical Cancer Research*. 15 de enero de 2014;20(2):511-21.
16. Dieci MV, Prat A, Tagliafico E, Paré L, Ficarra G, Bisagni G, et al. Integrated evaluation of PAM50 subtypes and immune modulation of pCR in HER2-positive breast cancer patients treated with chemotherapy and HER2-targeted agents in the CherLOB trial. *Annals of Oncology*. octubre de 2016;27(10):1867-73.
17. Carey LA, Berry DA, Cirincione CT, Barry WT, Pitcher BN, Harris LN, et al. Molecular Heterogeneity and Response to Neoadjuvant Human Epidermal Growth Factor Receptor 2 Targeting in CALGB 40601, a Randomized Phase III Trial of Paclitaxel Plus Trastuzumab With or Without Lapatinib. *J Clin Oncol*. 20 de febrero de 2016;34(6):542-9.
18. Llombart-Cussac A, Cortés J, Paré L, Galván P, Bermejo B, Martínez N, et al. HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. *The Lancet Oncology*. abril de 2017;18(4):545-54.
19. Swain SM, Ewer MS, Viale G, Delaloge S, Ferrero J-M, Verrill M, et al. Pertuzumab, trastuzumab, and standard anthracycline- and taxane-based chemotherapy for the neoadjuvant treatment of patients with HER2-positive localized breast cancer (BERENICE): a phase II, open-label, multicenter, multinational cardiac safety study. *Annals of Oncology*. marzo de 2018;29(3):646-53.
20. Prat A, Galván P, Jimenez B, Buckingham W, Jeiranian HA, Schaper C, et al. Prediction of Response to Neoadjuvant Chemotherapy Using Core Needle Biopsy Samples with the Prosigna Assay. *Clin Cancer Res*. 1 de febrero de 2016;22(3):560-6.

21. Alba E, Rueda OM, Lluch A, Albanell J, Chin S-F, Chacon JI, et al. Integrative cluster classification to predict pathological complete response to neoadjuvant chemotherapy in early breast cancer. *JCO*. 20 de mayo de 2018;36(15_suppl):579-579.
22. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. octubre de 2012;490(7418):61-70.
23. Zabala M, Lobo NA, Antony J, Heitink LS, Gulati GS, Lam J, et al. LEFTY1 Is a Dual-SMAD Inhibitor that Promotes Mammary Progenitor Growth and Tumorigenesis. *Cell Stem Cell*. agosto de 2020;27(2):284-299.e8.
24. Hancock BA, Chen Y-H, Solzak JP, Ahmad MN, Wedge DC, Brinza D, et al. Profiling molecular regulators of recurrence in chemorefractory triple-negative breast cancers. *Breast Cancer Res*. diciembre de 2019;21(1):87.
25. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. *Nature*. agosto de 2017;548(7667):297-303.
26. Mahner S, Baasch C, Schwarz J, Hein S, Wölber L, Jänicke F, et al. C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. *Br J Cancer*. 21 de octubre de 2008;99(8):1269-75.
27. Lu C, Shen Q, DuPré E, Kim H, Hilsenbeck S, Brown PH. cFos is critical for MCF-7 breast cancer cell growth. *Oncogene*. septiembre de 2005;24(43):6516-24.
28. Shi R, Peng H, Yuan X, Zhang X, Zhang Y, Fan D, et al. Down-regulation of c-fos by shRNA sensitizes adriamycin-resistant MCF-7/ADR cells to chemotherapeutic agents via P-glycoprotein inhibition and apoptosis augmentation. *Journal of Cellular Biochemistry*. 2013;114(8):1890-900.

29. Wu H, Bi J, Peng Y, Huo L, Yu X, Yang Z, et al. Nuclear receptor NR4A1 is a tumor suppressor down-regulated in triple-negative breast cancer. *Oncotarget*. 29 de abril de 2017;8(33):54364-77.
30. Zhou F, Drabsch Y, Dekker TJA, de Vinuesa AG, Li Y, Hawinkels LJAC, et al. Nuclear receptor NR4A1 promotes breast cancer invasion and metastasis by activating TGF- β signalling. *Nature Communications* [Internet]. diciembre de 2014 [citado 5 de marzo de 2019];5(1). Disponible en: <http://www.nature.com/articles/ncomms4388>
31. Fedorova O, Petukhov A, Daks A, Shuvalov O, Leonova T, Vasileva E, et al. Orphan receptor NR4A3 is a novel target of p53 that contributes to apoptosis. *Oncogene*. marzo de 2019;38(12):2108-22.
32. Klintman M, Buus R, Cheang MCU, Sheri A, Smith IE, Dowsett M. Changes in Expression of Genes Representing Key Biologic Processes after Neoadjuvant Chemotherapy in Breast Cancer, and Prognostic Implications in Residual Disease. *Clin Cancer Res*. 15 de mayo de 2016;22(10):2405-16.
33. Byrne CE, Decombe J-B, Bingham GC, Remont J, Miller LG, Khalif L, et al. Evaluation of Extracellular Matrix Composition to Improve Breast Cancer Modeling. *Tissue Engineering Part A*. 1 de abril de 2021;27(7-8):500-11.
34. Xiang L, Song Z, Rong G. Taxotere-induced WNT16 Expression in Carcinoma-Associated Fibroblasts Might Associate with Progression and Chemoresistance of Breast Cancer. *Ann Clin Lab Sci*. 3 de enero de 2020;50(2):205-12.
35. Kim J-H, Kim T-W, Kim S-J. Downregulation of ARFGEF1 and CAMK2B by promoter hypermethylation in breast cancer cells. *BMB Reports*. 31 de agosto de 2011;44(8):523-8.
36. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality

- Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209–49.
37. Elston CW, Ellis IO. pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19(5):403–10.
 38. Duffy MJ, Harbeck N, Nap M, Molina R, Nicolini A, Senkus E, et al. Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *European Journal of Cancer*. abril de 2017;75:284–98.
 39. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. agosto de 2000;406(6797):747–52.
 40. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*. 11 de septiembre de 2001;98(19):10869–74.
 41. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*. 15 de septiembre de 2015;5(10):2929–43.
 42. Yersal O. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *WJCO*. 2014;5(3):412.
 43. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Journal of Clinical Oncology* [Internet]. 9 de febrero de 2009 [citado 4 de octubre de 2019]; Disponible en: <https://ascopubs.org/doi/pdf/10.1200/JCO.2008.18.1370>
 44. Paik S, Kim C, Baehner FL, Park T, Wickerham DL, Wolmark N. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *The New England Journal of Medicine*. 2004;10.

45. van de Vijver MJ, He YD, van 't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *N Engl J Med.* 19 de diciembre de 2002;347(25):1999–2009.
46. Filipits M, Rudas M, Jakesz R, Dubsy P, Fitzal F, Singer CF, et al. A New Molecular Predictor of Distant Recurrence in ER-Positive, HER2-Negative Breast Cancer Adds Independent Information to Conventional Clinical Risk Factors. *Clin Cancer Res.* 15 de septiembre de 2011;17(18):6012–20.
47. Zaheed M, Wilcken N, Willson ML, O'Connell DL, Goodwin A. Sequencing of anthracyclines and taxanes in neoadjuvant and adjuvant therapy for early breast cancer. *Cochrane Database of Systematic Reviews [Internet].* 2019 [citado 2 de agosto de 2019]; Disponible en: <https://www.readcube.com/articles/10.1002/14651858.CD012873.pub2>
48. Ogston KN, Miller ID, Payne S, Hutcheon AW, Sarkar TK, Smith I, et al. A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. *The Breast.* 1 de octubre de 2003;12(5):320–7.
49. Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of Residual Breast Cancer Burden to Predict Survival After Neoadjuvant Chemotherapy. *Journal of Clinical Oncology.* octubre de 2007;25(28):4414–22.
50. Diana A, Carlino F, Franzese E, Oikonomidou O, Criscitiello C, De Vita F, et al. Early Triple Negative Breast Cancer: Conventional Treatment and Emerging Therapeutic Landscapes. *Cancers.* 29 de marzo de 2020;12(4):819.
51. Goutsouliak K, Veeraraghavan J, Sethunath V, De Angelis C, Osborne CK, Rimawi MF, et al. Towards personalized treatment for early stage HER2-positive breast cancer. *Nat Rev Clin Oncol.* abril de 2020;17(4):233–50.

52. Goorts B, van Nijnatten TJA, de Munck L, Moosdorff M, Heuts EM, de Boer M, et al. Clinical tumor stage is the most important predictor of pathological complete response rate after neoadjuvant chemotherapy in breast cancer patients. *Breast Cancer Research and Treatment*. mayo de 2017;163(1):83–91.
53. Bownes RJ, Turnbull AK, Martinez–Perez C, Cameron DA, Sims AH, Oikonomidou O. On–treatment biomarkers can improve prediction of response to neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res*. diciembre de 2019;21(1):73.
54. Esserman LJ, Berry DA, DeMichele A, Carey L, Davis SE, Buxton M, et al. Pathologic Complete Response Predicts Recurrence–Free Survival More Effectively by Cancer Subset: Results From the I–SPY 1 TRIAL—CALGB 150007/150012, ACRIN 6657. *J Clin Oncol*. 10 de septiembre de 2012;30(26):3242–9.
55. Gass P, Lux MP, Rauh C, Hein A, Bani MR, Fiessler C, et al. Prediction of pathological complete response and prognosis in patients with neoadjuvant treatment for triple–negative breast cancer. *BMC Cancer*. diciembre de 2018;18(1):1051.
56. Jarząb M, Stobiecka E, Badora–Rybicka A, Chmielik E, Kowalska M, Bal W, et al. Association of breast cancer grade with response to neoadjuvant chemotherapy assessed postoperatively. *pjp*. 2019;70(2):91–9.
57. Penault–Llorca F, Radošević–Robin N. Biomarkers of residual disease after neoadjuvant therapy for breast cancer. *Nature Reviews Clinical Oncology*. agosto de 2016;13(8):487–503.
58. Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, et al. Pharmacogenomic Predictor of Sensitivity to Preoperative Chemotherapy With Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide in Breast Cancer. *JCO*. 10 de septiembre de 2006;24(26):4236–44.

59. Liedtke C, Hatzis C, Symmans WF, Desmedt C, Haibe-Kains B, Valero V, et al. Genomic Grade Index Is Associated With Response to Chemotherapy in Patients With Breast Cancer. *JCO*. 1 de julio de 2009;27(19):3185–91.
60. Iddawela M, Rueda O, Eremin J, Eremin O, Cowley J, Earl HM, et al. Integrative analysis of copy number and gene expression in breast cancer using formalin-fixed paraffin-embedded core biopsy tissue: a feasibility study. *BMC Genomics*. diciembre de 2017;18(1):526.
61. Di Cosimo S, Triulzi T, Pizzamiglio S, De Cecco L, de Azambuja E, Fumagalli D, et al. The 41-gene classifier TRAR predicts response of HER2 positive breast cancer patients in the NeoALTTO study. *European Journal of Cancer*. septiembre de 2019;118:1–9.
62. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 1 de julio de 2011;121(7):2750–67.
63. Masuda H, Baggerly KA, Wang Y, Zhang Y, Gonzalez-Angulo AM, Meric-Bernstam F, et al. Differential Response to Neoadjuvant Chemotherapy Among 7 Triple-Negative Breast Cancer Molecular Subtypes. *Clinical Cancer Research*. 1 de octubre de 2013;19(19):5533–40.
64. Santonja A, Sánchez-Muñoz A, Lluch A, Chica-Parrado MR, Albanell J, Chacón JI, et al. Triple negative breast cancer subtypes and pathologic complete response rate to neoadjuvant chemotherapy. *Oncotarget* [Internet]. 29 de mayo de 2018 [citado 16 de septiembre de 2019];9(41). Disponible en: <http://www.oncotarget.com/fulltext/25413>
65. Jones RL, Salter J, A'Hern R, Nerurkar A, Parton M, Reis-Filho JS, et al. The prognostic significance of Ki67 before and after neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat*. julio de 2009;116(1):53–68.

66. Minckwitz G von, Schmitt WD, Loibl S, Müller BM, Blohmer JU, Sinn BV, et al. Ki67 Measured after Neoadjuvant Chemotherapy for Primary Breast Cancer. *Clin Cancer Res.* 15 de agosto de 2013;19(16):4521–31.
67. Takahashi Y, Iwamoto T, Suzuki Y, Kajiwara Y, Hatono M, Tsukioki T, et al. Evaluation of Therapeutic Target Gene Expression Based on Residual Cancer Burden Classification After Neoadjuvant Chemotherapy for HER2–Negative Breast Cancer. *Clinical Breast Cancer.* septiembre de 2019;S1526820919306433.
68. Luen SJ, Salgado R, Dieci MV, Vingiani A, Curigliano G, Gould RE, et al. Prognostic implications of residual disease tumor–infiltrating lymphocytes and residual cancer burden in triple–negative breast cancer patients after neoadjuvant chemotherapy. *Annals of Oncology.* febrero de 2019;30(2):236–42.
69. Sheri A, Smith IE, Johnston SR, A’Hern R, Nerurkar A, Jones RL, et al. Residual proliferative cancer burden to predict long–term outcome following neoadjuvant chemotherapy. *Ann Oncol.* enero de 2015;26(1):75–80.
70. Mittendorf EA, Wu Y, Scaltriti M, Meric–Bernstam F, Hunt KK, Dawood S, et al. Loss of HER2 Amplification Following Trastuzumab–Based Neoadjuvant Systemic Therapy and Survival Outcomes. *Clin Cancer Res.* 1 de diciembre de 2009;15(23):7381–8.
71. Guarneri V, Dieci MV, Barbieri E, Piacentini F, Omarini C, Ficarra G, et al. Loss of HER2 positivity and prognosis after neoadjuvant therapy in HER2–positive breast cancer patients. *Annals of Oncology.* diciembre de 2013;24(12):2990–4.
72. Xian Z, Quinones AK, Tozbikian G, Zynger DL. Breast cancer biomarkers before and after neoadjuvant chemotherapy: does repeat testing impact therapeutic management? *Human Pathology.* abril de 2017;62:215–21.
73. Korde LA, Lusa L, McShane L, Lebowitz PF, Lukes L, Camphausen K, et al. Gene expression pathway analysis to predict response to neoadjuvant docetaxel and

- capecitabine for breast cancer. *Breast Cancer Res Treat.* febrero de 2010;119(3):685–99.
74. I-SPY 1 TRIAL Investigators, Magbanua MJM, Wolf DM, Yau C, Davis SE, Crothers J, et al. Serial expression analysis of breast tumors during neoadjuvant chemotherapy reveals changes in cell cycle and immune pathways associated with recurrence and response. *Breast Cancer Res.* diciembre de 2015;17(1):73.
75. Balko JM, Giltane JM, Wang K, Schwarz LJ, Young CD, Cook RS, et al. Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. *Cancer Discov.* febrero de 2014;4(2):232–45.
76. Pelizzari G, Gerratana L, Basile D, Fanotto V, Bartoletti M, Liguori A, et al. Post-neoadjuvant strategies in breast cancer: From risk assessment to treatment escalation. *Cancer Treatment Reviews.* 1 de enero de 2019;72:7–14.
77. Salony, Solé X, Alves CP, Dey-Guha I, Ritsma L, Boukhali M, et al. AKT Inhibition Promotes Nonautonomous Cancer Cell Survival. *Mol Cancer Ther.* 1 de enero de 2016;15(1):142–53.
78. Balko JM, Cook RS, Vaught DB, Kuba MG, Miller TW, Bhola NE, et al. Profiling of residual breast cancers after neoadjuvant chemotherapy identifies DUSP4 deficiency as a mechanism of drug resistance. *Nat Med.* julio de 2012;18(7):1052–9.
79. Balko JM, Schwarz LJ, Bhola NE, Kurupi R, Owens P, Miller TW, et al. Activation of MAPK pathways due to DUSP4 loss promotes cancer stem cell-like phenotypes in basal-like breast cancer. *Cancer Res.* 15 de octubre de 2013;73(20):6346–58.
80. Bhola NE, Balko JM, Dugger TC, Kuba MG, Sánchez V, Sanders M, et al. TGF- β inhibition enhances chemotherapy action against triple-negative breast cancer [Internet]. 2013 [citado 4 de octubre de 2019]. Disponible en: <https://www.jci.org/articles/view/65416/pdf>

81. Masuda N, Lee S-J, Ohtani S, Im Y-H, Lee E-S, Yokota I, et al. Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N Engl J Med.* junio de 2017;376(22):2147-59.
82. Pinto JA, Araujo J, Cardenas NK, Morante Z, Doimi F, Vidaurre T, et al. A prognostic signature based on three-genes expression in triple-negative breast tumours with residual disease. *NPJ Genom Med.* 3 de febrero de 2016;1:15015.
83. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *J Clin Oncol.* 1 de junio de 2010;28(16):2784-95.
84. Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. *Arch Pathol Lab Med.* 2014;138:16.
85. Ribelles N, Jerez JM, Urda D, Subirats JL, Márquez A, Quero C, et al. Galén: Sistema de Información para la gestión y coordinación de procesos en un servicio de Oncología. 2010;6:12.
86. Molania R, Gagnon-Bartsch JA, Dobrovic A, Speed TP. A new normalization for Nanostring nCounter gene expression data. *Nucleic Acids Research.* 9 de julio de 2019;47(12):6073-83.
87. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M, et al. Evaluation of a 30-Gene Paclitaxel, Fluorouracil, Doxorubicin, and Cyclophosphamide Chemotherapy Response Predictor in a Multicenter Randomized Trial in Breast Cancer. *Clin Cancer Res.* 1 de noviembre de 2010;16(21):5351-61.

88. Hatzis C. A Genomic Predictor of Response and Survival Following Taxane–Anthracycline Chemotherapy for Invasive Breast Cancer. *JAMA*. 11 de mayo de 2011;305(18):1873.
89. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 1 de agosto de 2014;30(15):2114–20.
90. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 1 de marzo de 2010;26(5):589–95.
91. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 15 de agosto de 2009;25(16):2078–9.
92. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 3 de enero de 2012;22(3):568–76.
93. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *The Lancet*. julio de 2014;384(9938):164–72.
94. Mahadevappa R, Neves H, Yuen SM, Bai Y, McCrudden CM, Yuen HF, et al. The prognostic significance of Cdc6 and Cdt1 in breast cancer. *Sci Rep*. diciembre de 2017;7(1):985.
95. Deng Y, Jiang L, Wang Y, Xi Q, Zhong J, Liu J, et al. High expression of CDC6 is associated with accelerated cell proliferation and poor prognosis of epithelial ovarian cancer. *Pathology – Research and Practice*. abril de 2016;212(4):239–46.
96. Horne GA, Stewart HJS, Dickson J, Knapp S, Ramsahoye B, Chevassut T. Nanog Requires BRD4 to Maintain Murine Embryonic Stem Cell Pluripotency and Is Suppressed by Bromodomain Inhibitor JQ1 Together with Lefty1. *Stem Cells and Development*. abril de 2015;24(7):879–91.

97. Lobo NA, Shimono Y, Qian D, Clarke MF. The Biology of Cancer Stem Cells. Annual Review of Cell and Developmental Biology. 2007;23(1):675–99.
98. Babak Nami, Zhixiang Wang. HER2 in Breast Cancer Stemness: A Negative Feedback Loop towards Trastuzumab Resistance. Cancers. 26 de abril de 2017;9(12):40.
99. Valkenburg KC, Graveel CR, Zylstra–Diegel CR, Zhong Z, Williams BO. Wnt/ β –catenin Signaling in Normal and Cancer Stem Cells. Cancers (Basel). 19 de abril de 2011;3(2):2050–79.
100. Sun J, Cui H, Gao Y, Pan Y, Zhou K, Huang J, et al. TGF- α Overexpression in Breast Cancer Bone Metastasis and Primary Lesions and TGF- α Enhancement of Expression of Pro-cancer Metastasis Cytokines in Bone Marrow Mesenchymal Stem Cells. BioMed Research International. 2018;2018:1–10.
101. Schafer JM, Lehmann BD, Gonzalez–Ericsson PI, Marshall CB, Beeler JS, Redman LN, et al. Targeting MYCN–expressing triple–negative breast cancer with BET and MEK inhibitors. Sci Transl Med [Internet]. 11 de marzo de 2020 [citado 30 de marzo de 2021];12(534). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7427123/>
102. Bergaggio E, Piva R. Wild–Type IDH Enzymes as Actionable Targets for Cancer Therapy. Cancers. 19 de abril de 2019;11(4):563.
103. Alkaf A, Al–Jafari A, Wani TA, Alqattan S, Zargar S. Expression of STK11 gene and its promoter activity in MCF control and cancer cells. 3 Biotech [Internet]. diciembre de 2017 [citado 30 de marzo de 2021];7(6). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5628056/>
104. Uehiro N, Sato F, Pu F, Tanaka S, Kawashima M, Kawaguchi K, et al. Circulating cell–free DNA–based epigenetic assay can detect early breast cancer. Breast Cancer Res. diciembre de 2016;18(1):129.

105. Helsten T, Elkin S, Arthur E, Tomson BN, Carter J, Kurzrock R. The FGFR Landscape in Cancer: Analysis of 4,853 Tumors by Next-Generation Sequencing. *Clin Cancer Res.* 1 de enero de 2016;22(1):259–67.
106. Levine KM, Priedigkeit N, Basudan A, Tasdemir N, Sikora MJ, Sokol ES, et al. FGFR4 overexpression and hotspot mutations in metastatic ER+ breast cancer are enriched in the lobular subtype. *NPJ Breast Cancer* [Internet]. 27 de junio de 2019 [citado 8 de noviembre de 2019];5. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6597581/>
107. Garcia-Recio S, Thennavan A, East MP, Parker JS, Cejalvo JM, Garay JP, et al. FGFR4 regulates tumor subtype differentiation in luminal breast cancer and metastatic disease. *J Clin Invest.* 130(9):4871–87.
108. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 18 de abril de 2012;486(7403):346–52.
109. Yu K-D, Zhu R, Zhan M, Rodriguez AA, Yang W, Wong S, et al. Identification of Prognosis-Relevant Subgroups in Patients with Chemoresistant Triple-Negative Breast Cancer. *Clin Cancer Res.* 15 de mayo de 2013;19(10):2723–33.
110. Amplification of a calcium channel subunit CACNG4 increases breast cancer metastasis | Elsevier Enhanced Reader [Internet]. [citado 20 de abril de 2021]. Disponible en: <https://reader.elsevier.com/reader/sd/pii/S2352396420300219?token=20C82E1E795375580DD66E4E998A5A53B4F77F1421184F30CB1D6E66EA1735C5043837E99A1A89840B81F11BE31EC153&originRegion=eu-west-1&originCreation=20210420054402>
111. Fang Y, Zhang C, Wu T, Wang Q, Liu J, Dai P. Transcriptome Sequencing Reveals Key Pathways and Genes Associated with Cisplatin Resistance in Lung Adenocarcinoma A549 Cells. *PLOS ONE.* 2017;17.

112. Dethlefsen C, Højfeldt G, Hojman P. The role of intratumoral and systemic IL-6 in breast cancer. *Breast Cancer Res Treat.* 2013;8.
113. Iliopoulos D, Hirsch HA, Wang G, Struhl K. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proceedings of the National Academy of Sciences.* 25 de enero de 2011;108(4):1397-402.
114. Korkaya H, Liu S, Wicha MS. Breast cancer stem cells, cytokine networks, and the tumor microenvironment. *J Clin Invest.* 3 de octubre de 2011;121(10):3804-9.
115. Wu Y-H, Huang Y-F, Chang T-H, Chou C-Y. Activation of TWIST1 by COL11A1 promotes chemoresistance and inhibits apoptosis in ovarian cancer cells by modulating NF- κ B-mediated IKK β expression. *International Journal of Cancer.* 2017;141(11):2305-17.
116. Wu Y-H, Chang T-H, Huang Y-F, Chen C-C, Chou C-Y. COL11A1 confers chemoresistance on ovarian cancer cells through the activation of Akt/c/EBP β pathway and PDK1 stabilization. *Oncotarget.* 10 de junio de 2015;6(27):23748-63.
117. Wu Y-H, Chang T-H, Huang Y-F, Huang H-D, Chou C-Y. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene.* junio de 2014;33(26):3432-40.
118. Yu X, Li S, Pang M, Du Y, Xu T, Bai T, et al. TSPAN7 Exerts Anti-Tumor Effects in Bladder Cancer Through the PTEN/PI3K/AKT Pathway. *Front Oncol [Internet].* 8 de enero de 2021 [citado 26 de abril de 2021];10. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7821430/>
119. Xu Y, Bismar TA, Su J, Xu B, Kristiansen G, Varga Z, et al. Filamin A regulates focal adhesion disassembly and suppresses breast cancer cell migration and invasion. 207:17.

120. Ding L. Genome remodelling in a basal-like breast cancer metastasis and xenograft. 2010;464:7.
121. Moore KM, Thomas GJ, Duffy SW, Warwick J, Gabe R, Chou P, et al. Therapeutic Targeting of Integrin $\alpha\beta6$ in Breast Cancer. JNCI: Journal of the National Cancer Institute [Internet]. agosto de 2014 [citado 26 de abril de 2021];106(8). Disponible en: <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/dju169>
122. Yu F, Li J, Chen H, Fu J, Ray S, Huang S, et al. Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. :12.
123. Akaogi K. KLF4 suppresses estrogen-dependent breast cancer growth by inhibiting the transcriptional activity of ER α ; :9.
124. Nagata T. KLF4 and NANOG are prognostic biomarkers for triple-negative breast cancer. Breast Cancer. 2017;10.
125. Ray D, Yun YC, Idris M, Cheng S, Boot A, Iain TBH, et al. A tumor-associated splice-isoform of *MAP2K7* drives dedifferentiation in MBNL1-low cancers via JNK activation. Proc Natl Acad Sci USA. 14 de julio de 2020;117(28):16391-400.
126. Goetz MP, Kalari KR, Suman VJ, Moyer AM, Yu J, Visscher DW, et al. Tumor Sequencing and Patient-Derived Xenografts in the Neoadjuvant Treatment of Breast Cancer. JNCI: Journal of the National Cancer Institute [Internet]. julio de 2017 [citado 26 de febrero de 2021];109(7). Disponible en: <http://academic.oup.com/jnci/article/doi/10.1093/jnci/djw306/3064536/Tumor-Sequencing-and-PatientDerived-Xenografts-in>
127. Spencer ES, Bluemn EG, Johnston R, Zhang X, Gordon RR, Lewinshtein D, et al. Association of decreased expression of protein phosphatase 2A subunit PR55y (PPP2R2C) with an increased risk of metastases and prostate cancer-specific mortality. JCO. 20 de mayo de 2012;30(15_suppl):4669-4669.

128. Phan NN, Wang C-Y, Chen C-F, Sun Z, Lai M-D, Lin Y-C. Voltage-gated calcium channels: Novel targets for cancer therapy. *Oncol Lett.* agosto de 2017;14(2):2059-74.
129. Bieche I. Expression of PEA3/E1AF/ETV4, an Ets-related transcription factor, in breast tumors: positive links to MMP2, NRG1 and CGB expression. *Carcinogenesis.* 4 de diciembre de 2003;25(3):405-11.
130. Dumortier M, Ladam F, Damour I, Vacher S, Bièche I, Marchand N, et al. ETV4 transcription factor and MMP13 metalloprotease are interplaying actors of breast tumorigenesis. *Breast Cancer Res.* diciembre de 2018;20(1):73.
131. Yuan Z-Y, Dai T, Wang S-S, Peng R-J, Li X-H, Qin T, et al. Overexpression of ETV4 protein in triple-negative breast cancer is associated with a higher risk of distant metastasis. *Onco Targets Ther.* 26 de septiembre de 2014;7:1733-42.
132. Lim MJ, Wang XW. Nucleophosmin and human cancer. *Cancer Detect Prev.* 2006;30(6):481-90.
133. Yip SP, Siu PM, Leung PHM, Zhao Y, Yung BYM. The Multifunctional Nucleolar Protein Nucleophosmin/NPM/B23 and the Nucleoplasmin Family of Proteins. *The Nucleolus.* 23 de mayo de 2011;15:213-52.
134. Benci JL, Xu B, Qiu Y, Wu TJ, Dada H, Twyman-Saint Victor C, et al. Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. *Cell.* diciembre de 2016;167(6):1540-1554.e12.



UNIVERSIDAD
DE MÁLAGA



UNIVERSIDAD
DE MÁLAGA