UNIVERSITY OF VERONA

DEPARTMENT OF

Diagnostic and Public Health

GRADUATE SCHOOL OF

Health and Life Sciences

DOCTORAL PROGRAM IN

Inflammation, Immunity and Cancer

XXXII / 2016

TITLE OF THE DOCTORAL THESIS

IDENTIFICATION OF CLINICAL, PATHOLOGICAL AND MOLECULAR PREDICTORS OF PROGNOSIS IN EARLY-STAGE INVASIVE LOBULAR BREAST CANCER: ROLE OF PROLIFERATION AND CDK4/6 PATHWAY

S.S.D. MED/06

Tutor: Prof. Michele Milella

Coordinatore: Prof.ssa Gabriela Constantin

Dottorando: Dott.ssa Luisa Carbognin

TABLE OF CONTENTS

SUMMARY……3

ABSTRACT……5

INTRODUCTION……..7

MATERIAL AND METHODS

Prognostic role of Ki67 for early-stage ILC and comparison of the optimal Ki67 cut-off for ILC with the optimal cut-off for IDC……..….9

Investigation of the impact on survival of adjuvant chemotherapy in the lobular histotype……...…….....11

Building and validation of a prognostic nomogram for early-stage ILC patients in order to identify those prognostic 'outliers' candidate to undergo to genomic analysis……...……...12

RESULTS

Prognostic role of Ki67 for early-stage ILC and comparison of the optimal Ki67 cut-off for ILC with the optimal cut-off for IDC…….….18

Investigation of the impact on survival of adjuvant chemotherapy in the lobular histotype……..…….....26

Building and validation of a prognostic nomogram for early-stage ILC patients in order to identify those prognostic 'outliers' candidate to undergo to genomic analysis……...……...29

DISCUSSION……...…….42

CONCLUSION……..…...45

REFERENCES……..…...46

SUMMARY

Diversamente dal carcinoma duttale infiltrante (IDC) della mammella, per il carcinoma lobulare infiltrante (ILC) non sono ancora completamente noti quali siano i fattori patologici e molecolari che guidano la prognosi di questo istotipo. A questo proposito, è necessario chiarire l'eventuale impatto prognostico e/o predittivo di potenziali *drivers*, al fine di studiare il background molecolare dei pazienti con ILC. Tra i fattori da esplorare, quello della proliferazione, comunemente valutata mediante l'antigene Ki67, rappresenta ad oggi un aspetto rilevante nella scelta del trattamento per il tumore della mammella in fase iniziale, in particolare nella malattia luminale. L'analisi condotta su una casistica retrospettiva multicentrica di 679 pazienti con istologia lobulare, in stadio iniziale e operati, confrontata con 418 pazienti con istologia duttale ha evidenziato che un valore del 4%-5% rappresenta il miglior cut-off di Ki67 in grado di distinguere tra pazienti affetti da ILC a prognosi favorevole o sfavorevole.

Per quanto riguarda il ruolo della chemioterapia adiuvante ad oggi le principali linee guida non raccomandano un trattamento diverso in funzione dell'istologia, data l'assenza di studi randomizzati condotti nel ILC. Da analisi retrospettive emergono dati discordanti in merito al beneficio della chemioterapia adiuvante in aggiunta alla terapia ormonale per il ILC ad immunofenotipo luminale. A questo proposito, una differenza significativa in termini di sopravvivenza globale (OS) tra sola ormonoterapia e ormonoterapia più chemioterapia adiuvante (OS a 5 e 10-anni 96.3% vs 86.0% e 92.2% vs 67.5%, rispettivamente) è stata evidenziata dalla nostra analisi condotta mediante 'propensity score' su 473 pazienti affetti da ILC luminale.

Una delle principali strategie di ricerca emergenti in ambito oncologico si basa sullo studio del genoma di pazienti con risposta anomala al trattamento o con prognosi inattesa. Adottando questa strategia, abbiamo analizzato retrospettivamente una serie multicentrica di quasi 500 pazienti affetti da ILC sottoposti a resezione chirurgica e abbiamo costruito uno dei primi modelli di classificazione del rischio per ILC, successivamente convalidato in una coorte di 282 pazienti. Questo modello, basato su una combinazione di parametri semplici clinico-patologici, è stato in grado di stratificare in modo efficace i pazienti con ILC con una buona accuratezza prognostica. Una volta identificati i *'migliori'* e *'peggiori'* dal punto di

vista prognostico, abbiamo studiato il loro pattern molecolare, principalmente mediante sequenziamento genetico (NGS), e il loro profilo di espressione per identificare alterazioni molecolari ricorrenti ed esplorare la loro associazione prognostica in una coorte preliminare di 20 pazienti a buona prognosi e 14 a prognosi sfavorevole. In generale, il gene mutato più comunemente era il gene *CDH1* (38,2%), seguito da *PIK3CA* (29,4%) e *TP53* (20,6%), mentre la perdita di *CDH1* (44,1%) e *ARID1A* (38,2%) sono state le variazioni del numero di copie più frequenti. Le alterazioni molecolari erano distribuite indipendentemente dalla prognosi, ad eccezione dell'aumento del numero di copie (*gain*) di *CDK4*, esclusivamente presente nel sottogruppo a prognosi sfavorevole (35.0%, *p=0.03*; Odds Ratio 7.98, 95%CI 1.51-42.1, *p=0.014*).

Lo scopo ultimo del progetto era di valutare il profilo molecolare degli ILC resecabili utilizzando tecnologie moderne al fine di identificare quelle aberrazioni molecolari potenzialmente in grado di prevedere la probabilità di recidiva. Questa analisi integrata e multi-step eseguita ha suggerito che la via di CDK4/6 possa avere un impatto biologico rilevante sull'oncogenesi del ILC. Certamente, il carattere dei nostri dati, anche in considerazione della numerosità del campione, rappresenta un'ipotesi che deve essere prospetticamente validata in una serie di pazienti più estesa.

ABSTRACT

Differently from invasive ductal carcinoma (IDC) of the breast, pathological and molecular factors that guide the prognosis of invasive lobular carcinoma (ILC) are not completely known. In this regard, the prognostic and/or predictive impact of potential *drivers* needs to be elucidated, in order to create a global portrait for ILC patients. Among these factors, the role of proliferation, commonly evaluated by the Ki67 antigen, represents a relevant aspect in the choice of adjuvant treatment for breast cancer, in particular in the luminal setting. This retrospective multicentric analysis, including 679 patients with early-stage resected lobular histology and comparing these with 418 patients affected by IDC, showed that a value of 4%-5% represents the best Ki67 cut-off of able to significantly discriminate the prognosis of patients affected by ILC.

Regarding the role of adjuvant chemotherapy, to date the main guidelines do not recommend a different treatment approach according to histology, given the absence of randomized studies conducted in the ILC setting. From retrospective analyses, the benefit of adjuvant chemotherapy in addition to hormonotherapy for luminal ILC is still unclear. In this regard, our propensity score analysis conducted on 473 patients affected by luminal ILC showed a significant difference in terms of overall survival between hormonotherapy alone and hormonotherapy plus adjuvant chemotherapy (5- and 10-year 96.3% vs. 86.0% and 92.2% vs. 67.5%, respectively).

Nowadays, one of the main emerging research strategies in cancer is based on the study of the genome of exceptional responder and prognostic '*outlier*' patients. Adopting this strategy, we retrospectively analysed a multicenter series of nearly 500 ILC patients underwent surgical resection and we built one of the first risk classification model for ILC, subsequently validated in a cohort of 282 patients.

This model, based on a combination of simple and easily available clinical-pathological parameters, was able to effectively stratify ILC patients in prognostic risk classes, with a good accuracy. Once identified the 'best' and 'worst' prognostic performers, we investigated their molecular portrait, principally by next-generation sequencing (NGS), and their expression profile to identify recurrent molecular alterations and explore their association with prognosis, in a preliminary cohort of

20 patients with good prognosis and 14 with poor prognosis. Overall, the most frequent mutated gene was the *CDH1* gene (38.2%), followed by *PIK3CA* (29.4%) and *TP53* (20.6%), while the loss of *CDH1* (44.1%) and *ARID1A* (38.2%) were the most frequent copy number variation events. The molecular alterations were distributed regardless of the prognosis, except for the gain of *CDK4*, exclusively present in the poor prognosis subgroup (35.0%, *p=0.03*; Odds Ratio 7.98, 95%CI 1.51-42.1, *p=0.014*).

The final aim of the overall project was to evaluate the molecular profile of outliers resected ILC using modern technologies in order to identify those molecular aberrations that could potentially predict the probability of recurrence. This integrated and multi-step analysis suggested that the CDK4/6 pathway may have a significant biological impact on the ILC prognosis. Certainly, this hypothesis needs to be prospectively validated in a larger series.

INTRODUCTION

Invasive lobular carcinoma (ILC), accounting for 5% to 15% of all invasive breast tumors, represents the second most common histologic type of breast cancer after invasive ductal carcinoma (IDC) [1, 2].

The lobular histotype is generally characterized by luminal/HER2-negative phenotype and lower mitotic index, usually measured by the immunohistochemical assessment of the Ki67 antigen, and grading in comparison with IDC. Although these aspects are typically associated with a favorable prognosis, the overall long-term survival of ILC appears worse than that of invasive carcinoma of NST [3]. Even the recent genomic characterization of ILC underlined that this histotype represents a peculiar entity of breast cancer [4-6].

However, despite the several dissimilarities between lobular and ductal histology in terms of clinical-pathological and molecular features, the prognostic and therapeutic aspects of ILC are currently borrowed from the ductal histotype, thereby limiting their clinical utility in the specific context of lobular subtype. In the adjuvant context, patients resected for early-stage breast cancer are assigned to receive hormonotherapy or chemotherapy according to international guidelines based upon immunophenotype, clinical-pathological and genomic features, regardless of the histotype [7]. Therefore, the identification of prognostic factors for ILC represents a relevant aspect for clinical practice in order to select appropriate treatment strategies.

The preliminary stratification of patients according to prognosis may allow to identify those patients characterized by ‘best’ and ‘worst’ prognosis, emerging as ’outliers’, or rather biologically different from the majority of the population affected by the same biologically-defined disease. These ‘outlier’ patients may harbor a series of molecular aberrations potentially driving their peculiar clinical behavior. In this regard, the assignment of a reliable clinical significance to a specific genomic alteration, that can impact on patient prognosis and determine susceptibility to selective targeted therapies, represents a main challenge for recent translational research. This approach to selectively identify potential predictors of prognosis and (eventually) resistance or response to a given treatment in the context of ‘best’ and ‘worst’ prognostic performers, represents nowadays one of the strategy

that may successfully integrate the clinical findings with the advanced genetic acquisitions [8].

In the recent era of personalized medicine, the identification of the appropriate risk category for each patient represents a promising strategy for two main reasons [8]. First, in the context of an early-stage disease, the prognostic stratification might allow selection of those patients with a more favourable risk-benefit ratio from adjuvant treatments. Second, from an exploratory point-of-view, the molecular characterization of patients featured by a different prognosis, by applying the modern technologies, could help in the identification of those genomic and epigenomic aberrations potentially able to predict the probability of disease recurrence (prognostic factors) and the efficacy of agents selectively targeting these candidate pathways (predictive factors). In this regard, the project aims to: 1) explore the prognostic role of Ki67 in the context of early-stage ILC and compare the optimal Ki67 cut-off for ILC with the optimal cut-off for IDC; 2) investigate the effect of adjuvant chemotherapy in the lobular histotype; 3) develop and validate a prognostic nomogram for early stage ILC patients, according to the combination of clinical-pathological predictors, in order to identify those prognostic 'outliers' candidate to undergo to genomic analysis.

MATERIAL AND METHODS

The overall project included three main aims:

1. Exploring the prognostic role of Ki67 for early-stage ILC and comparing the optimal Ki67 cut-off for ILC with the optimal cut-off for IDC.
2. Investigation of the impact on survival of adjuvant chemotherapy in the lobular histotype.
3. Building and validation of a clinical prognostic nomogram for early-stage ILC patients, according to the combination of clinical-pathological predictors, in order to identify those prognostic 'outliers' candidate to undergo to genomic analysis.

1. Prognostic role of Ki67 for early-stage ILC and comparison of the optimal Ki67 cut-off for ILC with the optimal cut-off for IDC.

Patients' Population.

Data of consecutive patients affected by early stage 'pure' ILC, undergone surgery at 3 Italian institutions (University of Verona and University of Padua providing data for the training set [TS] and Regina Elena National Cancer Institute (Rome) for the validation set [VS]) from January 1990 and December 2013, were considered eligible. These data were compared with clinical-pathological data of a consecutive series of patients affected by pure IDC and undergone surgery between 2006 and 2010. Inclusion criteria were 'pure' ILC or IDC diagnosis (stage I-III), curative surgery and availability of clinical-pathological parameters (age, Performance Status, menopausal status, type of surgery, clinical stage, treatments, grading, Ki67, estrogen receptor, progesterone receptor, and HER2 status). The study was approved by the local Ethics Committee (Prot. CESC n° 24163, May 20th, 2014).

End-Points.

The aims of this analysis were: 1) to identify the best prognostic cut-off of ki67 in terms of disease free survival (DFS) and overall survival (OS) for patients with resected ILC, comparing that with the best prognostic cut-off for patients with

resected IDC; 2) to investigate its impact in long-term outcome; 3) to validate the Ki67 cut-off for ILC in an external patients' cohort. The DFS was defined by the time between diagnosis and local or distant recurrence, onset of secondary cancer or death for any cause and OS was defined by the time between diagnosis and death for any cause.

Statistics.

Descriptive statistic was adopted to summarize pertinent study information. Follow-up was analyzed and reported according to Shuster [9]. The maximally selected Log-Rank statistics analysis was applied to the Ki67 continuous variable in order to estimate the most appropriate cut-off values able to split patients with ILC and IDC diagnosis into groups with different DFS probabilities [10]. Associations between variables and groups according to Ki67 were analyzed (Chi-square test). Clinical-pathological data were correlated to DFS and OS using a Cox model. The hazard ratio (HR) and the 95% Confidence interval (95% CI) were estimated using the Cox univariate model [11]. A multivariate Cox proportional hazard model was developed using stepwise regression (forward selection, enter limit and remove limit, $p=0.10$ and $p=0.15$, respectively), to identify independent predictors of outcomes in patients with ILC and IDC. The Harrell's guidelines for the identification of the correct number of covariates were taken into account for the power analysis [12]. Outcomes calculated by the Kaplan–Meier product limit method. The log-rank test was used to assess differences between subgroups. Significance was defined at $p<0.05$.

To address the multivariate model overfit, a cross-validation technique, which evaluates the replication stability of the final Cox model in predicting the outcomes, was investigated [13, 14].

The Subpopulation Treatment Effect Pattern Plot (STEPP) analysis was performed to evaluate whether the prognostic effect (in terms of DFS) between ILC and IDC varies according to the Ki67 [15].

Finally, an external validation of the Ki67 cut-off for ILC patients was accomplished as well: the maximally selected Log-Rank statistics analysis was applied to the Ki67 values in the VS.

The SPSS® (18.0), R® (2.6.1), and MedCalc® (14.2.1) licensed statistical programs were used for all analyses.

2. Investigation of the impact on survival of adjuvant chemotherapy in the lobular histotype.

Patients' Population.

Data of consecutive patients affected by early stage 'pure' ILC, undergone surgery at 3 Italian institutions (University of Verona, University of Padua and Catholic University of the Sacred Heart Rome) between January 2000 and December 2013, were collected. Inclusion criteria were 'pure' ILC diagnosis (stage I-III), curative surgery and availability of clinical-pathological parameters (age, Performance Status, menopausal status, type of surgery, clinical stage, treatments, grading, Ki67, estrogen receptor, progesterone receptor, and HER2 status).

End-Point.

The aim of this analysis was to investigate the impact of the addition of adjuvant chemotherapy to hormonotherapy in patients affected by luminal early-stage pure ILC in terms of DFS, distant disease-free survival (DDFS) and OS. The DDFS was defined by the time between diagnosis and distant recurrence.

Statistics.

Descriptive statistic was adopted to summarize pertinent study information. Follow-up was analyzed and reported according to Shuster [9]. Clinical-pathological data were correlated to DFS, DDFS and OS using a Cox model. The hazard ratio (HR) and the 95% Confidence interval (95% CI) were estimated using the Cox univariate model [11]. A multivariate Cox proportional hazard model was developed using stepwise regression (forward selection, enter limit and remove limit, p=0.10 and p=0.15, respectively), to identify independent predictors of outcomes. The Harrell's guidelines for the identification of the correct number of covariates were taken into account for the power analysis [12]. A propensity score analysis was performed to evaluate the prognostic impact of adjuvant chemotherapy

[16]. Outcomes calculated by the Kaplan–Meier product limit method. The log-rank test was used to assess differences between subgroups. Significance was defined at $p<0.05$. The SPSS® (18.0), R® (2.6.1), and MedCalc® (14.2.1) licensed statistical programs were used for all analyses.

3. Building and validation of a prognostic nomogram for early-stage ILC patients in order to identify those prognostic 'outliers' candidate to undergo to genomic analysis.

Patients' Population.

Clinical charts of consecutive patients affected by early stage ILC diagnosed at 3 Italian institutions (University of Verona and University of Padova providing data for the TS and University Federico II of Napoli for the VS) between January 1990 to December 2013 were considered eligible. Inclusion criteria were 'pure' ILC diagnosis (stage I-III), curative surgery and availability of clinical-pathological parameters.

End-Points.

The aims of this analysis were 1) to develop a prognostic nomogram on the basis of clinico-pathological factors in a multi-center population of ILC (TS), in order to identify prognostic 'outliers'; 2) to validate the model in an external patients' cohort (VS); 3) to explore potential molecular drivers of prognosis with NGS in a subset of prognostic 'outlier' patients.

Statistical Analysis.

Descriptive statistics was used to summarize pertinent study information. Follow-up was analyzed and reported according to Shuster et al [9]. Associations between variables were analyzed according to the Pearson $\chi 2$ test. The assessment of interactions between significant investigational variables was taken into account when developing the multivariate model. The Harrell's guidelines for the identification of the correct number of covariates were taken into account for the power analysis (the number of deaths should have been more than 10 times greater than the number of investigated predictors, so that the expected error from the Cox

model would be less than 10%). The Kaplan-Meier product limit method was adopted for survival analyses. The HR and the 95% CI were estimated for each variable using the Cox univariate model [11]. The variables considered at univariate analysis included age, performance status, menopausal status (the postmenopausal status was defined as the absence of a menstrual period > 12 months, due to natural causes, or bilateral oophorectomy), type of surgery, clinical stage (tumor (T)-category and nodal status), grading, Ki67, vascular invasion, estrogen receptor, progesterone receptor, and HER2 status. A multivariate hazard model was developed using the stepwise regression (forward selection, enter limit and remove limit, p=0.10 and p=0.15, respectively), to identify independent predictors of outcomes [17]. The outcomes were DFS and OS. To address the multivariate model overfit, an internal cross-validation technique was performed [12-14, 18]. The internal validation analysis generates a number of simulation datasets (at least 100, each approximately 80% of the original size), by randomly selecting patients from the original sample, to establish the consistency of the model across less-powered patient' samples. The log-rank test was used to assess differences between subgroups. Significance was defined at the $p<0.05$ level. The SPSS® (18.0), R® (2.6.1), and MedCalc® (14.2.1) licensed statistical programs were used for all analyses.

Prognostic Score Assessment.

A step-by-step protocol was followed according to the methodological approach for building a nomogram for cancer prognosis proposed by *Iasonos et al*. [14] with respect to the reporting recommendations for tumor marker prognostic studies (REMARK) criteria [19, 20]. The log-HR obtained from the Cox multivariate analysis was used to derive weighting factors of a continuous prognostic index, aimed to identify differential outcomes' risks. Coefficients estimates were 'normalized' dividing by the smallest one and rounding the resulting ratios to the nearest integer value. Thus, a continuous score assigning to patients an 'individualized' risk was generated. To develop the prognostic model [21], patients' outcomes were displayed by dividing patients into three risk classes, by considering cut-offs chosen at approximately equal distance along the range of values [22].

Finally, an external validation of the DFS model was explored in the VS. The receiver operating characteristic analysis allowed to estimate the accuracy of the prognostic model, by the AUC) determination with SE.

Samples and molecular analysis.

Formalin-fixed and paraffin-embedded (FFPE) ILC samples from patients at poor prognosis (defined with the prognostic model) were selected from the University of Verona according to the availability of a tissue sample from the surgical specimen of the primary tumor. In order to compare the molecular pattern of these samples at poor prognosis with those at good, we selected a subgroup of cases at good prognosis according to the developed prognostic model. The collected material has been subjected to targeted NGS analysis for somatic mutation (SM), copy number variation (CNV) and transcriptomic analysis. In addition, quantitative-PCR, immunohistochemistry (IHC) and the fluorescent in situ hybridization (FISH) for the validation of gene alterations of interest were performed. Finally, stromal tumor infiltrating lymphocytes (sTIL) were also evaluated. Fisher's exact test corrected for multiple comparisons was used as appropriate. In addition, *Peto* odds ratios (OR) for estimating the risk of association of a given biomarker with each prognostic class was determined.

DNA extraction and qualification.

DNA was obtained from tumour and non-neoplastic tissue matched included in FFPE blocks. In particular, tumour DNA from FFPE was prepared after enrichment for neoplastic cellularity to at least 70% using manual microdissection of 10 consecutive 4-µm FFPE sections. Sections were then purified using the QIAamp DNA FFPE Tissue Kit (Qiagen) and qualified as reported elsewhere [23].

Mutational, CNV and transcriptomic analyses by targeted NGS.

Matched tumour/normal DNA and RNA from all FFPE samples was subjected to NGS. To analyse DNA, an Ion Ampliseq custom panel was used to investigated SM and CNV status of all exons of 26 selected genes upon the results of published WGS and exome data [4, 24]: AKT1, ARID1A, ATM, BRCA1, BRCA2, CCND1, CDH1, CDK4, CDKN2A, ERBB2, ESR1, FGFR1, FOXA1, GATA3, MAP3K1, MTOR,

MYC, PALB2, PDGFRA, PDL1, PGR, PIK3CA, PTEN, RB1, TBX3, and TP53. Twenty nanograms of DNA were used for custom panel multiplex PCR amplification. The quality of the obtained libraries was evaluated by the Agilent 2100 Bioanalyzer on-chip electrophoresis (Agilent Technologies). Emulsion PCR to construct the libraries of clonal sequences was performed with the Ion OneTouch™ OT2 System (Life Technologies). Sequencing was run on the Ion Proton (PI, Life Technologies) loaded with Ion PI Chip v2. Data analysis, including alignment to the hg19 human reference genome and variant calling, was done using the Torrent Suite Software v.5.0 (Life Technologies). Filtered variants were annotated using a custom pipeline based on vcflib (https://github.com/ekg/vcflib), SnpSift [25], the Variant Effect Predictor (VEP) software [26] and NCBI RefSeq database. Additionally, alignments were visually verified with the Integrative Genomics Viewer (IGV) v2.3 [27] to further confirm the presence of mutations identified by exome and targeted sequencing. The CNV was evaluated compared BAM files of single tumor sample to three normal BAM files using OncoCNV 6.8 version software [28]. The mutational load was detected by dividing the number of non-synonymous mutations for the number of coding bases (mega bases) analyzed by sequencing. Mutational load and chromosome integrity number (CIN) were also evaluated [29]. The CIN was evaluated for each sample dividing length of altered chromosomes to length of chromosome regions investigated. Three CIN categories was identified: under 0.2; equal 0.2; major 0.2.

Copy Number Variation validation by Quantitative-PCR (q-PCR).

The q-PCR analysis of copy numbers was applied to all cases for followed genes: ARID1A, CDK4, ESR1, MTOR, PDL1 and PTEN. The target and reference assays were purchased from Applied Biosystems. RNaseP was used as endogenous control for normalization of analyzed locus. The following assays were used: ARID1A (Hs06542243), CDK4 (Hs02225231), ESR1 (Hs04321628), MTOR (Hs00873941), PDL1 (Hs03707126), PTEN (Hs05217581) and RNaseP (part number 4403326). The experimental procedure recommended by the manufacturer (Applied Biosystems) was followed. Twenty nanogram genomic DNA was used in the Q-PCR reaction, and a negative control was analyzed in parallel. All q-PCR reactions

were run in quadruplicate in a 7900HT machine (Applied Biosystems) using standard cycling conditions of 10 min at 95°C, followed by 40 cycles of [95°C for 15 sec and at 60°C for 1 min]. Pooled normal FFPE DNA was used as calibrator and as normal.

Variant calling criteria for mutations and CNVs.

Tumor mutations identified by Variant Caller Software v.5.0 (Thermo Fisher) has screened as follow: i) filtering-out of germline mutations identified in matched normal sample sequenced; ii) filtering-in mutations with at least 20 reads with alteration and with frequency major of 10%; iii) filtering-out artefacts through manual visualization of mutation using Integrative Genomics Viewer (IGV) v2.3. CNV detection was performed using OncoCNV 6.8 version software and followed criteria: i) a p-value under 0.05; ii) a q-value under 0.05. In presence of not suitable values, an orthogonal cross-validation using q-PCR was performed. In this case, only concordant results between methodical were reported.

mRNA profiling.

Available RNA was converted in cDNA and subsequently subjected to analysis using Ion Ampliseq Transcriptome Human Gene Expression Kit [30]. The matrix of raw counts (from Ion Torrent) was used for differential expression analysis with DESeq2 Bioconductor package [30]. DESeq2 was used in combination with RUVSeq for normalization purposes [31]. A batch factor of variation was calculated from the expression of empirical control genes, that is least significantly DE genes based on a preliminary DE analysis performed with edgeR, and such factor was then added to the DESeq2 design formula [32]. Here we used 5000 least significant DE genes and we retained K=3 factors of variation. The package ComplexHeatmap was used to draw the heatmaps [33].

Immunohistochemistry, FISH and sTIL assessment.

Immunohistochemistry for PD1, PD-L1, phmTOR, CDK4 and CDK6 was performed on surgical specimens using 4 µm FFPE tissue. Fluorescent in situ hybridization for PD-L1, CDK4, ESR1 and mTOR was performed as well. The

following primary antibodies for the assessment were adopted: PD1 clone NAT, Rabbit monoclonal, ABCAM, dilution 1:100; PD-L1 clone E1L3Nok, Rabbit monoclonal, Cell Signaling, dilution 1:100; phmTOR, Rabbit polyclonal, Cell Signaling, dilution 1:100; CDK6, ABCAM, anti-Cdk6 antibody (EPR4515), ab124821, diluition 1: 200; CDK4, A304-225A-M, Betyl Laboratories, dilution 1:400. Expression was graded based on the intensity and the percentage of stained cells (0= no staining, 1 = weak, 2 = moderate, 3=strong). The following primary chromosomal probes for the cytogenetic assessment were adopted: locus 1p36 (ph-mTOR locus 9p13-link 9p24.1 (PD-L1); locus 12q14 (CDK4) and locus 6q25.1 (ESR1). For each case 100 non-overlapping nuclei were counted. On 100 nuclei, the number of signals per nucleus is calculated as the average of these values. On the basis of the average of the signals' numbers, wild-type cases were distinguished from cases with aberration in the copies' number of the gene (gains as >15% of at least three fluorescent spots or deletions as presence of single fluorescent signals in >40% of nuclei after correction with non-neoplastic nuclei to avoid artefactual nuclear truncation). Stromal TIL were assessed on hematoxylin and eosin (H&E)-stained sections according to the International TILs Working Group 2014 recommendations [34].

RESULTS

1. Prognostic role of Ki67 for early-stage ILC and comparison of the optimal Ki67 cut-off for ILC with the optimal cut-off for IDC.

Patients.

Data from 457 (TS) and 222 (VS) ILC patients and 418 with IDC undergone surgery were gathered (overall 1097 patients; patients features for TS and the IDC cohort are reported in Table 1). Median age at diagnosis was 61 years (years) [range 35-96 years] for ILC (TS) and 59 years (28-94 years) for IDC. In the TS ILC cohort, one hundred-twenty-two (26.7%) recurrences occurred at a median follow-up of 75 months (range 1-396 months). Median DFS was 175 months (95% CI 153-196), with a 5- and 10-year rate of 82.5% and 71.4%, respectively. Median OS was 213 months (95% CI 190-236), with a 5- and 10-year rate of 91.8% and 81.7%, respectively. In the IDC cohort, fifty-seven (13.6%) recurrences occurred at a median follow up of 75 months (range 1-122 months). Median DFS was not reached, with a 5- and 10-year rate of 90.4% and 68.5%, respectively. Median OS was not reached, with a 5- and 10-year rate of 95.8% and 81.6%, respectively.

Table 1. Clinical-pathological and therapeutic characteristics in patients with invasive lobular carcinoma (training set) and invasive ductal carcinoma.

Clinical-pathological characteristics	Subcategories	ILC [TS] Patients N (%)	IDC Patients N (%)
Menopausal status	Premenopausal	133 (29.3)	143 (34.2)
	Postmenopausal	322 (70.7)	275 (65.8)
Performance status (ECOG)	0	265 (58.0)	371 (88.8)
	1	19 (4.2)	24 (5.7)
	2	2 (0.4)	2 (0.5)
	Unknown	171 (37.4)	21 (5.0)
Grading	1	116 (25.4)	55 (13.2)
	2	151 (33.0)	223 (53.3)
	3	62 (13.6)	137 (32.8)
	Unknown	128 (28.0)	3 (0.7)
Oestrogen Receptor status	Positive	426 (93.2)	343 (82.1)
	Negative	17 (3.7)	74 (17.7)
	Unknown	14 (3.1)	1 (0.2)
Progesterone Receptor status	Positive	381 (83.4)	321 (76.8)
	Negative	49 (10.7)	92 (22.0)
	Unknown	27 (6.0)	5 (1.2)
HER2 status	Positive	28 (6.1)	80 (19.1)
	Negative	324 (70.9)	324 (77.5)
	Unknown	105 (23.0)	14 (3.4)

T descriptor according to TNM [7° Edition]	1 2 3 4 Unknown	257 (56.2) 136 (29.8) 37 (8.1) 22 (4.8) 5 (1.1)	278 (66.5) 129 (30.9) 1 (0.2) 5 (1.2) 5 (1.2)
Lymph-nodal status at diagnosis	Positive Negative Unknown	168 (36.8) 275 (60.2) 14 (3.0)	121 (28.9) 268 (64.2) 29 (6.9)
Vascular Invasion	Present Absent Unknown	80 (17.5) 263 (57.5) 114 (24.9)	165 (39.5) 238 (56.9) 15 (3.6)
Multifocality	Present Absent Unknown	82 (17.9) 352 (77.0) 23 (5.0)	81 (19.4) 336 (80.4) 1 (0.2)
Type of surgery	Tumorectomy Quadrantectomy Mastectomy	119 (26.1) 161 (35.2) 177 (38.7)	265 (63.4) 68 (16.3) 85 (20.3)
Lymph-node dissection	Yes No Unknown	294 (64.3) 160(35.0) 3 (0.7)	203 (48.6) 215 (51.4) 0 (0)
Adjuvant hormonal therapy	Yes No	404 (88.4) 53 (11.6)	327 (78.2) 91 (21.8)
Adjuvant chemotherapy	Yes No	184 (40.3) 273 (59.7)	197 (47.1) 221 (52.9)
Adjuvant Trastuzumab	Yes No	11 (2.4) 446 (97.6)	61 (14.6) 357 (85.4)
Adjuvant radiotherapy	Yes No Unknown	284 (62.1) 164 (35.9) 9 (1.9)	289 (69.1) 129 (30.9) 0 (0)

Legend Table 1. ILC, invasive lobular carcinoma; IDC, invasive ductal carcinoma; TS, training set; N, number; ECOG, Eastern Cooperative Oncology Group.

Ki67 distribution and maximally selected Log-Rank statistics analysis.

As shown in Figure 1 (Panel A), tumor proliferation was significantly affected by histology: a statistically significant lower distribution of Ki67 immunostaining was found in ILC compared to IDC patients ($p<0.0001$). In the ILC cohort (TS), the optimal cut-off (absolute peak) of Ki67 identified by the maximally selected Log-Rank statistics Analysis was 4% for DFS (Figure 1, Panel B). Patients' characteristics and their differences according to group identified by the cut-off are reported in Table 2. In the IDC cohort, the optimal cut-off of Ki67 was 14%.

Figure 1. Whiskers-box plot of Ki67 values (Panel A) in patients with invasive lobular carcinoma and invasive ductal carcinoma; Maximally selected Log-rank statistics analysis (Panel B) of disease-free survival according to Ki67 (%) in the invasive lobular carcinoma cohort (training set). p-value, log-rank analysis.

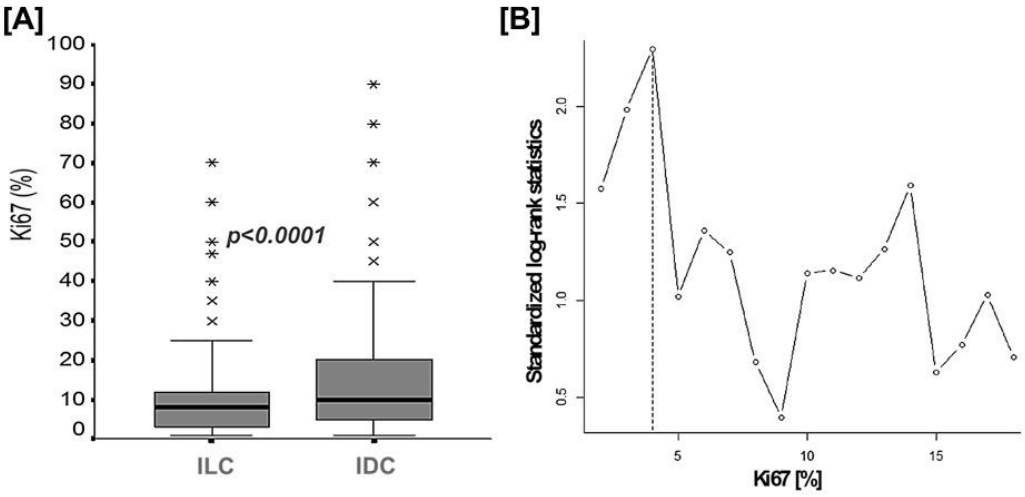


Table 2. Patients' characteristics according to Ki67 groups in patients with invasive lobular carcinoma (training set).

Clinical-pathological characteristics	Subcategories	Ki67 ≤4% [N=123] (%)	Ki67 >4% [N=296] (%)	*p-value*
Grading	1 2 3 Unknown	50 (40.7) 34 (27.6) 8 (6.5) 31 (25.2)	65 (22.0) 113 (38.2) 52 (17.6) 66 (22.2)	*<0.0001*
Oestrogen Receptor status	Positive Negative Unknown	116 (94.3) 7 (5.7) 0 (0)	285 (96.3) 9 (3.0) 2 (0.7)	*0.29*
Progesterone Receptor status	Positive Negative Unknown	100 (81.3) 18 (14.6) 5 (4.1)	259 (87.5) 30 (10.1) 7 (2.4)	*0.25*
HER2 status	Positive Negative Unknown	8 (6.5) 91 (74.0) 24 (19.5)	20 (6.8) 228 (77.0) 48 (16.2)	*0.72*
T descriptor according to TNM [7° Edition]	1 2 3 4 Unknown	81 (65.9) 31 (25.2) 9 (7.3) 1 (0.8) 1 (0.8)	159 (53.7) 92 (31.1) 27 (9.1) 18 (6.1) 0 (0)	*0.02*
Lymph-nodal status	Positive Negative Unknown	34 (27.6) 87 70.7) 2 (1.7)	121 (40.9) 165 (55.7) 10 (3.4)	*0.02*
Vascular Invasion	Present Absent Unknown	16 (13.0) 82 (66.7) 25 (20.3)	62 (20.9) 178 (60.1) 56 (19.0)	*0.16*
Multifocality	Present Absent Unknown	26 (21.1) 94 (76.4) 3 (2.4)	53 (17.9) 231 (78.0) 12 (4.1)	*0.56*

Type of surgery	Tumorectomy Quadrantectomy Mastectomy	39 (31.7) 46 (37.4) 38 (30.9)	76 (25.7) 103 (34.8) 117 (39.5)	*0.31*
Lymph-node dissection	Yes No Unknown	72 (58.5) 51 (41.5) 0 (0.0)	111 (37.5) 184 (62.2) 1 (0.3)	*0.62*
Adjuvant hormonal therapy	Yes No	109 (88.6) 14 (11.4)	276 (93.2) 20 (6.8)	*0.11*
Adjuvant chemotherapy	Yes No	43 (35.0) 80 (65.0)	129 (43.6) 167 (56.4)	*0.1*
Adjuvant radiotherapy	Yes No Unknown	85 (69.1) 34 (27.6) 4 (3.3)	106 (35.8) 185 (62.5) 5 (1.7)	*0.19*

Legend Table 2. N, number; p-value, chi-square test.

Multivariate Analysis and Internal Validation Analysis.

At the multivariate analysis for ILC patients, performance status and nodal status were independent predictors for DFS; performance status, nodal status, Ki67 and T-size were independent predictors of OS. At the internal cross-validation analysis, all factors were confirmed as independent factors for both DFS and OS (Table 3).

Table 3. Multivariate analysis in patients with invasive lobular carcinoma (training set).

ILC Predictors	**DFS HR (95% CI) [*p-value*]**	**Replication Rate [Internal Validation]**	**OS HR (95% CI) [*p-value*]**	**Replication Rate [Internal Validation]**
Performance Status (ECOG) [1-2 vs 0]	3.73 (1.50-9.30) [*0.005*]	100%	3.27 (1.49-7.18) [*0.003*]	96%
Nodal Status [Positive vs Negative]	3.75 (1.70-8.26) [*0.001*]	100%	2.96 (1.53-5.74) [*0.001*]	100%
Ki67 (>4% vs ≤4%)	-	-	2.28 (1.0-5.19) [*0.05*]	94%
T descriptor according to TNM [7° Edition] (3-4 vs 1-2)	-	-	2.6 (1.31-5.13) [*0.006*]	96%

Legend Table 3. ILC, invasive lobular carcinoma; DFS, disease-free survival; HR, Hazard Ratio; CI: confidence intervals; OS, overall survival, vs, versus; ECOG, Eastern Cooperative Oncology Group.

With regard to IDC patients, performance status, age at diagnosis, vascular invasion and grading were independent predictors of DFS; performance status, age at diagnosis, vascular invasion, T-size and hormonal receptor status were independent predictors of OS, with a high replication rate at the internal validation analysis

(Table 4).

Table 4. Multivariate analysis in patients with invasive ductal carcinoma.

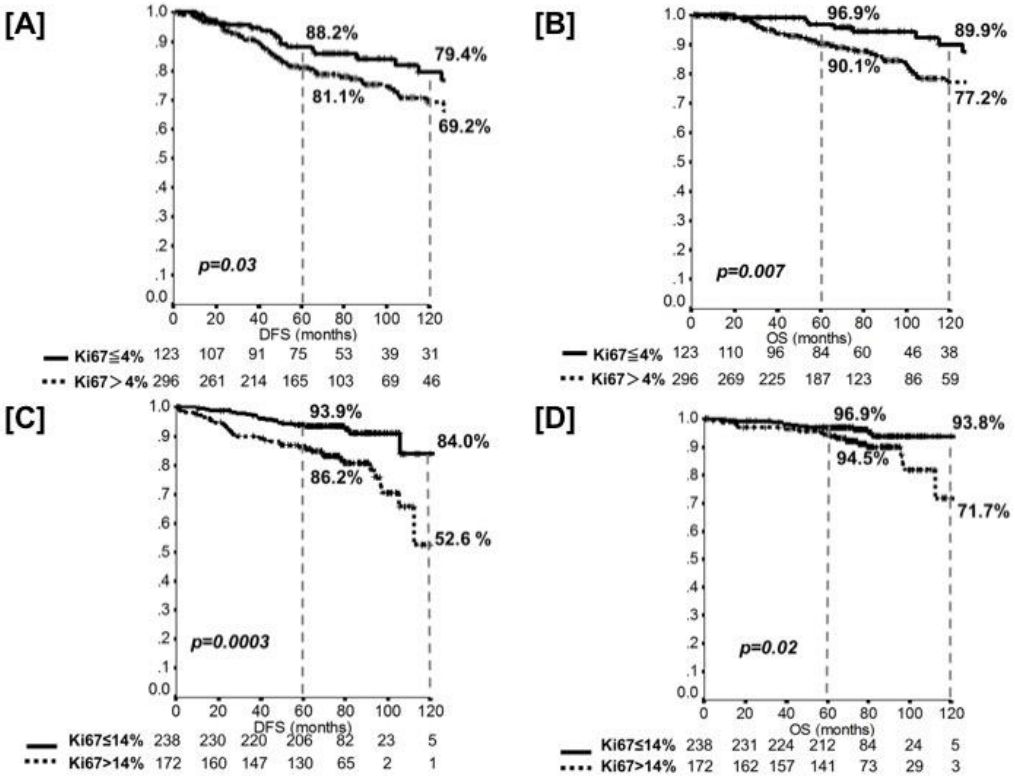
IDC Predictors	**DFS HR (95% CI) [*p-value*]**	**Replication Rate [Internal Validation]**	**OS HR (95% CI) [*p-value*]**	**Replication Rate [Internal Validation]**
Performance Status (ECOG) [1-2 vs 0]	3.39 (1.49-7.71) [*0.004*]	95%	5.64 (1.95-16.32) [*0.001*]	90%
Age at diagnosis [>60 vs ≤60]	2.40 (1.17-4.95) [*0.017*]	85%	8.68 (1.90-39.75) [*0.005*]	95%
Vascular Invasion [Present vs Absent]	2.01 (1.04-3.86) [*0.037*]	80%	2.37 (0.90-6.25) [*0.08*]	75%
Grading [3 vs 1-2]	3.60 (1.86-6.97) [*<0.0001*]	90%	-	-
T descriptor according to TNM [7° Edition] (3-4 vs 1-2)	-	-	6.29 (1.39-28.35) [*0.0.17*]	90%
Hormonal receptor Status [Neg. Vs Pos.]	-	-	8.07 (3.09-21.05) [*<0.001*]	90%

Legend Table 4. IDC, invasive ductal breast carcinoma; DFS, disease-free survival; HR, Hazard Ratio; CI: confidence intervals; OS, overall survival, vs, versus; ECOG, Eastern Cooperative Oncology Group; N, negative; Pos., positive.

Survival Analysis according to Ki67 cut-off.

In ILC patients, the optimal Ki67 cut-off was an independent predictor for OS, and it significantly discriminated the DFS. Survival curves according to Ki67 for ILC patients are shown in Figure 2, Panels A-B. In IDC patients, the optimal Ki67 cut-off was not independent at multivariate analysis; besides, Ki67 significantly discriminated the prognosis at unadjusted Kaplan-Meier curves (Figure 2, Panels C-D).

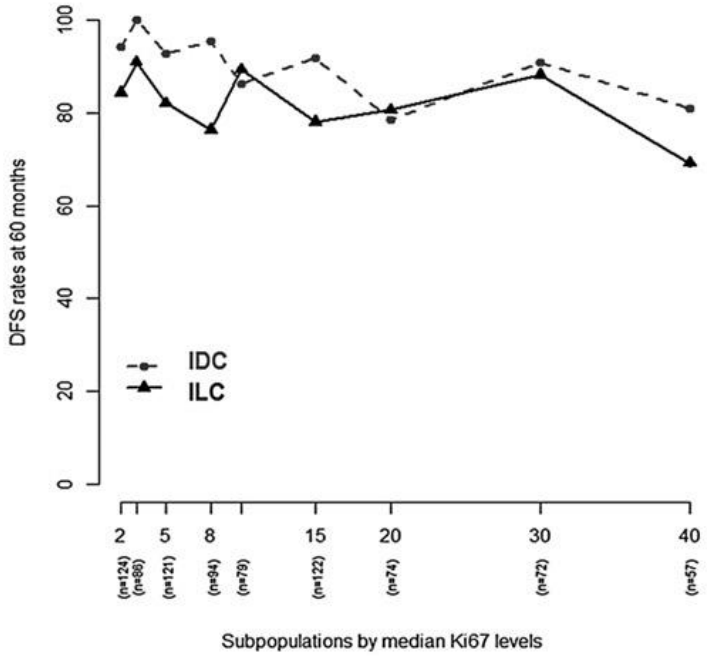
Figure 2. Disease-free survival (Panel A and C) and overall survival (Panel B and D) according to Ki67 in the invasive lobular carcinoma cohort (Panel A and B) and in the ductal invasive carcinoma cohort (Panel C and D). p-value, log-rank analysis; DFS, disease-free survival; OS, overall survival.



STEPP Analysis.

The STEPP analysis of Ki67 assay in terms of DFS rates at 60 months according to histology was reported in Figure 3. This analysis showed that in the presence of low values of Ki67, patients with IDC have a better DFS than patients with ILC, while with the increase of Ki67 value the prognosis tends to overlap.

Figure 3. Subpopulation Treatment Effect Pattern Plot (STEPP) analysis of Ki67 assay according to histology: DFS rates at 60 months of patients with invasive lobular and ductal carcinoma according to patients' subpopulations clustered by Ki67 (%). DFS, disease-free survival; ILC, invasive lobular carcinoma; IDC, invasive ductal carcinoma.



External Validation Analysis.

The VS consisted of 222 patients with operable or locally advanced ILC patients (Table 5). Median age was 59 years (range 33-88 years) and median follow-up was 71 months (range 1-195 months). The optimal Ki67 cut-off for OS, independent predictor at the multivariate analysis in the TS, was 5%. The found Ki67 cut-off significantly discriminated the prognosis in the VS (Figure 4).

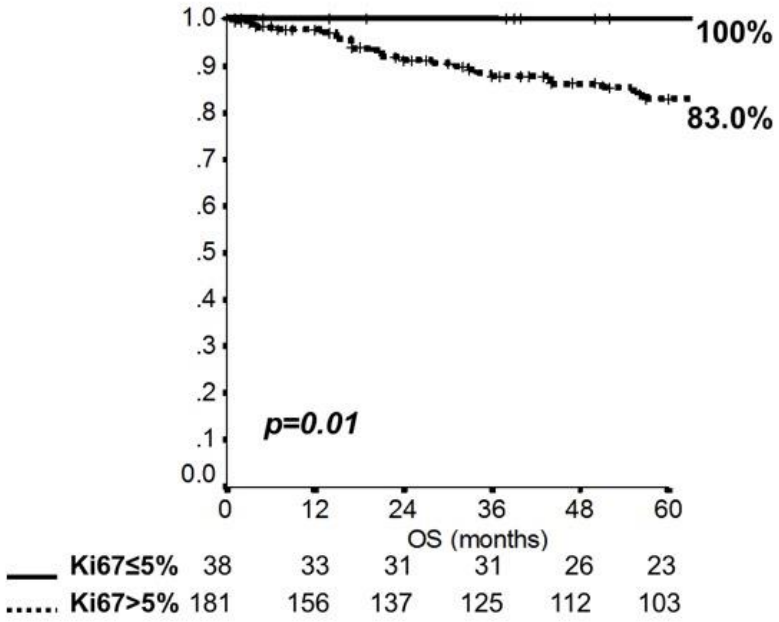
Table 5. Clinical-pathological and therapeutic characteristics in patients with invasive lobular carcinoma (validation set).

Characteristics	**Subcategories**	**ILC [VS] Patients N (%)**
Grading	1 2 3 Unknown	2 (0.9) 180 (81.1) 15 (6.8) 25 (11.3)
Oestrogen Receptor status	Positive Negative Unknown	184 (82.9) 38 (17.1) 0 (0.0)
Progesterone Receptor status	Positive Negative Unknown	161 (72.5) 61 (27.5) 0 (0.0)
HER2 status	Positive Negative Unknown	11 (5.0) 184 (82.9) 27 (12.2)

T descriptor according to TNM [7° Edition]	1 2 3 4 Unknown	113 (50.9) 81 (36.5) 15 (6.8) 7 (3.2) 6 (2.7)
Lymph-nodal status	Positive Negative Unknown	92 (41.4) 119 (53.6) 11 (5.0)
Adjuvant chemotherapy	Yes No	197 (47.1) 221 (52.9)

Legend Table 5. ILC, invasive lobular breast carcinoma; VS, validation set; N, number.

Figure 4. Overall survival according to Ki67 in the validation set of invasive lobular carcinoma. p-value, log-rank analysis; OS, overall survival.



2. Investigation of the impact on survival of adjuvant chemotherapy in the lobular histotype.

Patients.

Data from 473 patients with luminal/HER2-negative pure ILC were gathered (Table 6). Median age was 58 years [range 26-96]. At a median follow-up of 77 months [range 1-225], 403 pts (85.2%) were alive and 70 pts (14.8%) experienced locoregional or distant recurrence. The 5- and 10-yrs DFS rate were 84.6% and 72.9%, respectively. The 5- and 10-yrs rate OS were 92.1% and 80.7%, respectively. The 5- and 10-yrs rate DDFS were 88.5% and 77.9%, respectively.

Table 6. Clinical-pathological and therapeutic characteristics.

Characteristics	Subcategories	Patients N (%)
T descriptor according to TNM [7° Edition]	1 2 3 4 Unknown	268 (56.7) 143 (30.2) 35 (7.4) 16 (3.4) 11 (2.3)
Lymph-nodal status	0 1 2 3 Unknown	282 (59.6) 106 (22.4) 40 (8.5) 32 (6.8) 13 (2.7)
Ki67	≤4% >4% Unknown	100 (21.1) 334 (70.6) 39 (8.2)
Grading	1 2 3 Unknown	108 (22.8) 167 (35.3) 94 (19.9) 104 (22.0)
Type of Surgery	Tumorectomy Quadrantectomy Mastectomy	126 (26.6) 187 (39.5) 160 (33.8)
Adjuvant hormonal therapy	Yes No Unknown	415 (87.7) 29 (6.1) 29 (6.1)
Adjuvant chemotherapy	Yes No Unknown	209 (44.2) 235 (49.7) 29 (6.1)
Adjuvant radiotherapy	Yes No Unknown	244 (51.6) 191 (40.4) 38 (8.0)

Legend Table 6. N, number.

Multivariate Analysis.

Tumour-category according to TNM, lymph-node status and age at diagnosis were independent predictors for DFS at multivariate analysis. Nodal status and age were independent predictors for OS (Table 7).

Table 7. Independent predictors of outcome at multivariate analysis.

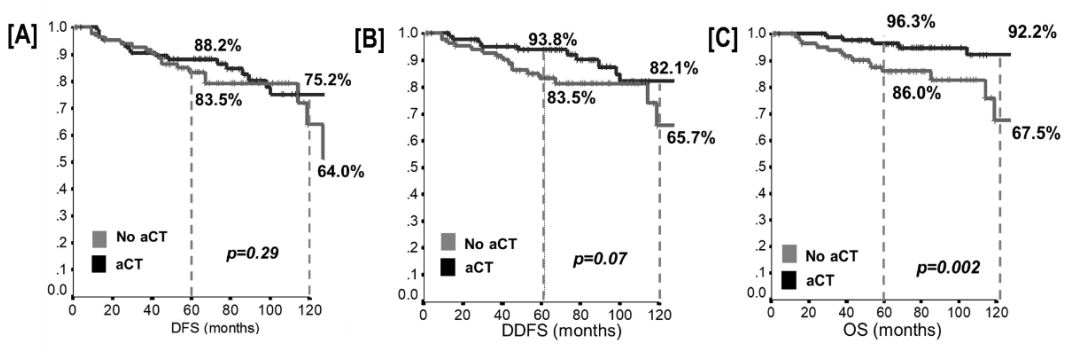
Predictors	**DFS HR (95% CI) [*p-value*]**	**OS HR (95% CI) [*p-value*]**
Age at diagnosis [>57 vs ≤57]	1.60 (0.96-2.67) *[p=0.072]*	2.71 (1.39-5.31) *[p=0.003]*
Nodal Status (Pos. vs Neg.)	2.62 (1.54-4.49) *[p<0.0001]*	4.46 (2.87-8.71) *[<0.0001]*
T-category according to TNM (2-4 vs 1)	1.31 (0.99-1.72) *[p=0.061]*	-

Legend Table 7. DFS, disease-free survival; HR, Hazard Ratio; CI: confidence intervals; OS, overall survival. vs, versus; Neg., negative; Pos., positive.

Propensity score analysis.

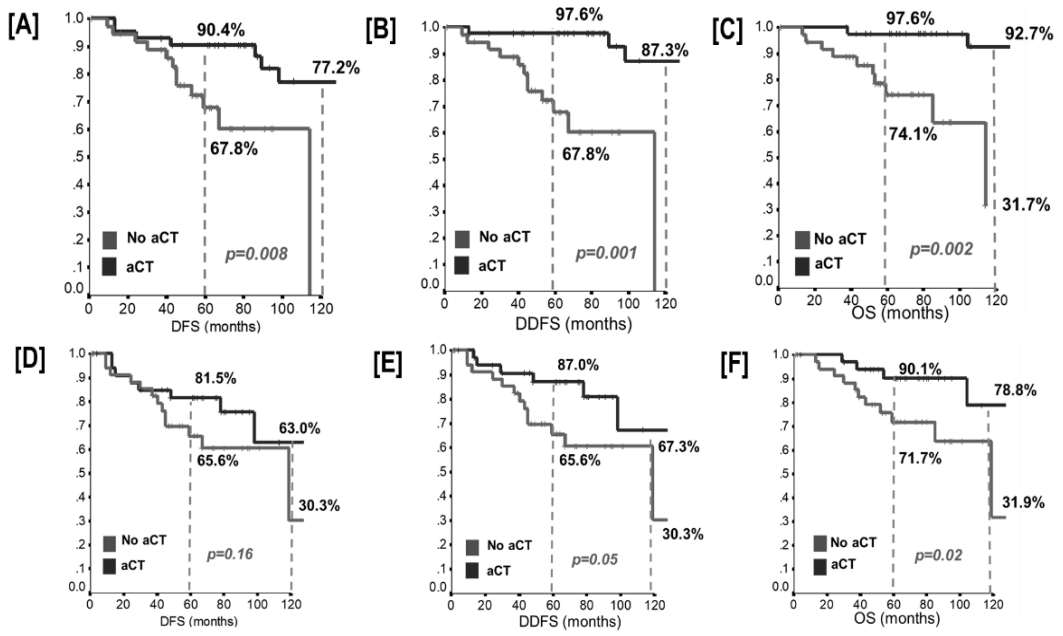
After adjusting for independent factors with the propensity score method (178 patients evaluable), no significant effect of adjuvant chemotherapy upon DFS and DDFS was found (Figure 5, Panels A-B), whereas a significant prognostic effect of the addition of adjuvant chemotherapy to adjuvant hormonotherapy upon OS was found (Figure 5, Panel C).

Figure 5. Disease-free survival (Panel A), distant disease-free survival (Panel B) and overall survival (Panel C) according to adjuvant chemotherapy. p-value, log-rank analysis; DFS, disease-free survival; DDFS, distant disease-free survival; OS, overall survival; aCT, adjuvant chemotherapy.



According to prognostic factors, adjuvant chemotherapy significantly prolongs DFS, DDFS and OS in patients with Tumor-category >1 and node-positive status (Figure 6). Moreover, adjuvant chemotherapy significantly prolongs OS in patients with Ki67>4% (*p=0.0009*) and Grading 2-3 (*p=0.01*).

Figure 6. Disease-free survival, distant disease-free survival and overall survival according to adjuvant chemotherapy in patients with Tumor-category <1 (Panels A-C) and in patients with node-positive status (Panels D-F). p-value, log-rank analysis; DFS, disease-free survival; DDFS, distant disease-free survival; OS, overall survival; aCT, adjuvant chemotherapy.



3. Building and validation of a prognostic nomogram for early-stage ILC patients in order to identify those prognostic 'outliers' candidate to undergo to genomic analysis.

Patients.

Data from overall 773 patients (491 patients for TS and 282 for VS) with ILC who underwent surgery were gathered. Patients' characteristics of the TS are listed in Table 8. At a median follow-up of 77 months (range 1-396), median DFS was 175 months (95% CI 153-197), with a 5-/10-year rate of 82.4%/70.5%, respectively. Median OS was 213 months (95% CI 190-236), with a 5-/10-year rate of 91.8%/82.2%, respectively.

Table 8. Clinico-pathological and therapeutic characteristics in patients with invasive lobular carcinoma (Training Set, N=491).

Characteristics	Subcategories	ILC [TS] Patients N (%)
Menopausal status	Premenopausal	142 (28.9)
	Postmenopausal	349 (71.1)
Grading	1	124 (25.3)
	2	155 (31.6)
	3	75 (15.3)
	Unknown	137 (27.9)
Oestrogen Receptor status	Positive	460 (93.7)
	Negative	17 (3.5)
	Unknown	14 (2.8)
Progesterone Receptor status	Positive	412 (83.9)
	Negative	52 (10.6)
	Unknown	27 (5.5)
Ki67	<5%	136 (27.7)
	≥5%	313 (63.7)
	Unknown	42 (8.6)
HER2 status	Positive	27 (5.5)
	Negative	353 (71.9)
	Unknown	111 (22.6)
T category according to TNM [7° Edition]	1	276 (56.2)
	2	148 (30.1)
	3	40 (8.1)
	4	23 (4.7)
	Unknown	4 (0.9)
Lymph-nodal status	Positive	180 (36.7)
	Negative	297 (60.5)
	Unknown	14 (2.8)
Vascular Invasion	Present	87 (17.7)
	Absent	290 (59.1)
	Unknown	114 (23.2)
Multifocality	Present	93 (18.9)
	Absent	375 (76.4)
	Unknown	23 (4.7)

Type of surgery	Tumorectomy Quadrantectomy Mastectomy	130 (26.5) 165 (33.6) 196 (39.9)
Adjuvant hormonal therapy	Yes No Unknown	432 (88.0) 534(11.0) 5 (1.0)
Adjuvant chemotherapy	Yes No	199 (40.5) 292 (59.5)
Adjuvant Trastuzumab	Yes No	11 (2.2) 480 (97.8)
Adjuvant radiotherapy	Yes No Unknown	301 (61.3) 177 (36.0) 13 (2.6)

Legend-Table 8. ILC, invasive lobular carcinoma; TS, training set; N, number.

Multivariate analysis.

At the multivariate analysis, T-category (1-2) and negative nodal status were independent predictors for longer DFS; age at diagnosis <60 years, negative nodal status and Ki67 <5% were independent predictors for longer OS. At the internal cross-validation analysis, all variables were confirmed as independent factors (Table 9).

Table 9. Multivariate analysis in patients with invasive lobular carcinoma (Training Set, N=491).

Predictors	**DFS** **HR (95% CI)** [*p-value*]	**Replication Rate** [Internal Validation]	**OS** **HR (95% CI)** [*p-value*]	**Replication Rate** [Internal Validation]
T-category according to TNM (7° Edition) [3-4 *vs.* 1-2]	1.78 (0.97-3.25) [*0.062*]	60%	-	-
Nodal Status [Positive *vs.* Negative]	2.46 (1.50-4.05) [*<0.0001*]	95%	3.32 (1.71-6.45) [*<0.0001*]	100%
Age [>60 years *vs.* ≤60 years]	-	-	2.19 (1.14-4.21) [*0.019*]	90%
Ki67 [≥5% *vs.* <5%]	-	-	2.47 (1.02-5.94) [*0.044*]	80%

Prognostic score assessment.

According to the HR obtained at the multivariate analysis, a prognostic scoring index was assigned to each patient to identify the individual risk of recurrence and death (Table 10). Based on the outcome, patients were divided into three risk classes for DFS and OS: low/intermediate/high risk of recurrence or death:

score<2/score=2/score>2. According to the prognostic model, a significant prognostic difference between patients at low, intermediate, and high risk was found for DFS (10-year: 76.3%, 67.6%, and 39.8%, *p<0.0001*; area under the curve [AUC] 0.60 (Standard Error [SE], 0.03)); Figure 7, Panel A) and OS (10-year: 92.7%, 82.7% and 67.1%, *p<0.0001;* AUC 0.66 (SE, 0.03)).

Table 10. Prognostic score assessment, according to outcome.

	Score points		
DFS	**0**	**1**	**2**
T-category (according to TNM 7° Edition)	1-2	3-4	-
Nodal status	Negative	-	Positive
OS			
Nodal status	Negative	-	Positive
Age at diagnosis	≤60 years	>60 years	-
Ki67	<5%	≥5%	-

Legend Table 10. DFS, disease-free survival; OS, overall survival.

External validation analysis.

The VS consisted of 282 patients (Table 11); median follow-up 86 months (range 1-348). The model, developed in the TS, has been proven to be equally able to discriminate the DFS in the VS (Figure 7, Panel B). Indeed, a significant prognostic difference between patients at low, intermediate, and high risk was found (10-year: 81.5%, 53.4%, 44.0%, p<0.0001; AUC 0.70 (SE, 0.03)). Considering the overall population (TS plus VS), the performance of the model adjusting for the set was confirmed (p<0.001), without difference in terms of DFS between the TS and VS (HR, 1.02 (95%CI 0.77-1.37, p=0.88).

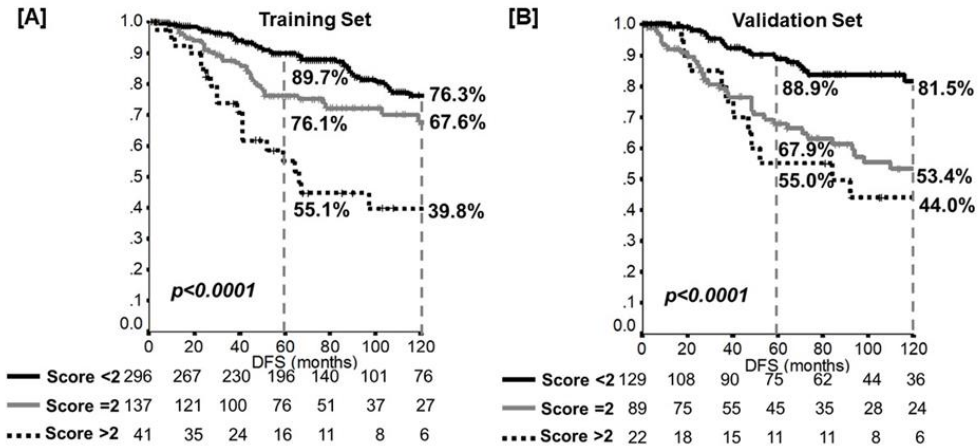
Table 11. Clinical-pathological and therapeutic characteristics in patients with invasive lobular carcinoma (Validation Set, N=282).

Characteristics	**Subcategories**	**ILC [VS]** **Patients N (%)**
Grading	1 2 3 Unknown	6 (2.1) 74 (26.2) 79 (28.0) 123 (43.6)
Oestrogen Receptor status	Positive Negative	234 (83.0) 12 (4.3)

	Unknown	36 (12.8)
Progesterone Receptor status	Positive Negative Unknown	215 (76.2) 30 (10.6) 37 (13.1)
Ki67	<5% ≥5% Unknown	8 (2.8) 158 (56.0) 116 (41.1)
HER2 status	Positive Negative Unknown	18 (6.4) 174 (61.7) 90 (31.9)
T category according to TNM [7° Edition]	1 2 3 4 Unknown	127 (45.0) 94 (33.3) 15 (5.3) 9 (3.2) 37 (13.1)
Lymph-nodal status	Positive Negative Unknown	127 (45.0) 140 (49.6) 15 (5.4)
Type of surgery	Tumorectomy Quadrantectomy Mastectomy	34 (12.1) 135 (47.9) 113 (40.1)
Adjuvant hormonal therapy	Yes No	212 (75.2) 70 (24.8)
Adjuvant chemotherapy	Yes No	200 (70.9) 82 (29.1)
Adjuvant Trastuzumab	Yes No	13 (4.6) 269 (95.4)

Legend Table 11. ILC, invasive lobular carcinoma; VS, validation set; N, number.

Figure 7. Disease-free survival (DFS) according to the risk-class model in the Training set [Panel A] and in the Validation set [Panel B]. p-value: log-rank analysis.



Patients' Sample.

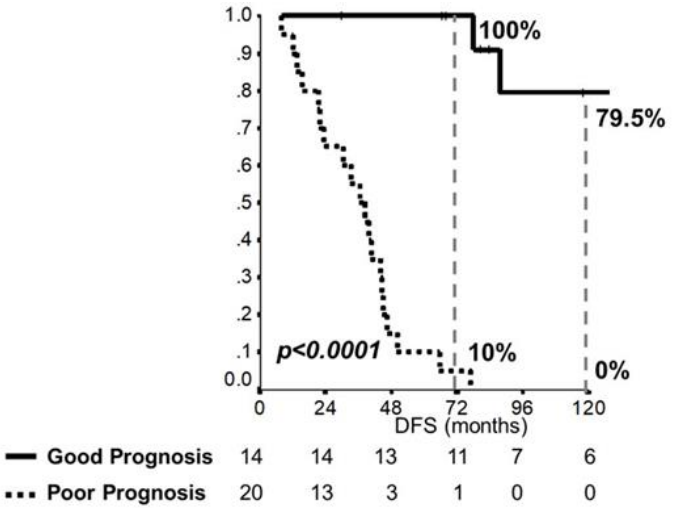
The tissue sample from the surgical specimen of the primary tumor matched with the clinico-pathological annotations, required by the inclusion criteria to be addressed to NGS analysis, was available at the coordinating center for 20 and 14 patients scored at poor and good prognosis class according to the DFS model, respectively (Table 12, Figure 8).

Table 12. Clinico-pathological and therapeutic characteristics in the 34 patients undergone molecular analysis according to prognosis.

Characteristics	Subcategories	Poor Group Patients N (%)	Good Group Patients N (%)
Grading	1	6 (30.0)	3 (21.4)
	2	10 (50.0)	9 (64.4)
	3	4 (20.0)	2 (14.2)
Oestrogen Receptor status	Positive	17 (85.0)	12 (85.8)
	Negative	1 (5.0)	1 (7.1)
	Unknown	2 (10.0)	1 (7.1)
Progesterone Receptor status	Positive	13 (65.0)	11 (78.6)
	Negative	3 (15.0)	0 (0)
	Unknown	4(20.0)	3 (21.4)
Ki67	<5%	2 (10.0)	8 (57.2)
	≥5%	17 (85.0)	5 (35.7)
	Unknown	1 (5.0)	1 (7.1)
HER2 status	Positive	2 (10.0)	1 (7.1)
	Negative	17 (85.0)	5 (35.7)
	Unknown	1 (5.0)	8 (57.2)
T category according to TNM [7° Edition]	1	3 (15.0)	8 (57.1)
	2	15 (75.0)	5 (35.7)
	3	1 (5.0)	1 (7.1)
	4	1 (5.0)	0 (0)
Lymph-nodal status	Positive	16 (80.0)	3 (21.4)
	Negative	3 (15.0)	11 (78.6)
	Unknown	1 (5.0)	0 (0)
Vascular Invasion	Present	16 (80.0)	3 (21.4)
	Absent	3 (15.0)	11 (78.6)
	Unknown	1 (5.0)	0 (0)
Adjuvant hormonal therapy	Yes	17 (85.0)	10 (71.4)
	No	3(15.0)	4 (28.6)
Adjuvant chemotherapy	Yes	11 (55.0)	8 (57.2)
	No	9 (45.0)	6 (42.8)
Adjuvant radiotherapy	Yes	12 (60.0)	9 (64.3)
	No	8 (40.0)	5 (35.7)

Legend-Table 12. N, number.

Figure 8. Disease-free survival (DFS) in patients at poor and good prognosis selected for molecular analysis. p-value: log-rank analysis.



Molecular features according to prognosis.

Mutations were observed in 29 cases of all series for 26 genes analysed. In detail: one mutation was observed in 10/34 cases (29.4%), more than one in 19/34 (55.9%) cases while no alteration in 5/34 cases (14.7%). A mean of 12.4 mutations/Mb was achieved. The most commonly mutated gene in the whole cohort was *CDH1* (38.2%), followed by *PIK3CA* (29.4%) and *TP53* (20.6%) (Figure 9 (Panel A), Table 13). Copy number variation was observed in all cases: one CNV was observed in 5/34 cases (14.7%), while more than one in 29/34 cases (85.3%). Loss of heterozygosis of *CDH1* (44.1%) and *ARID1A* (38.2%) were the most frequent CNV events, followed by gain in *FGFR1* and *ESR1* (each 12/34; 35.3%). The prevalence of gene somatic mutation (SM) and CNV according to prognostic groups was reported in Table 13. Interestingly, gain of *CDK4* was (7/34; 21.2%) exclusively present in this poor prognosis group, whereas no good prognosis case showed this alteration (*p=0.03*). Moreover, *CDK4* gain (at NGS analysis) resulted in a statistically significant higher chance to be associated with poor prognosis (OR 7.98, 95%CI 1.51-42.1, *p=0.014*) (Figure 9, Panel B).

Table 13. Prevalence of somatic mutations and copy number variations analysis of the 26 genes in the 34 invasive lobular carcinoma patients according to prognostic groups.

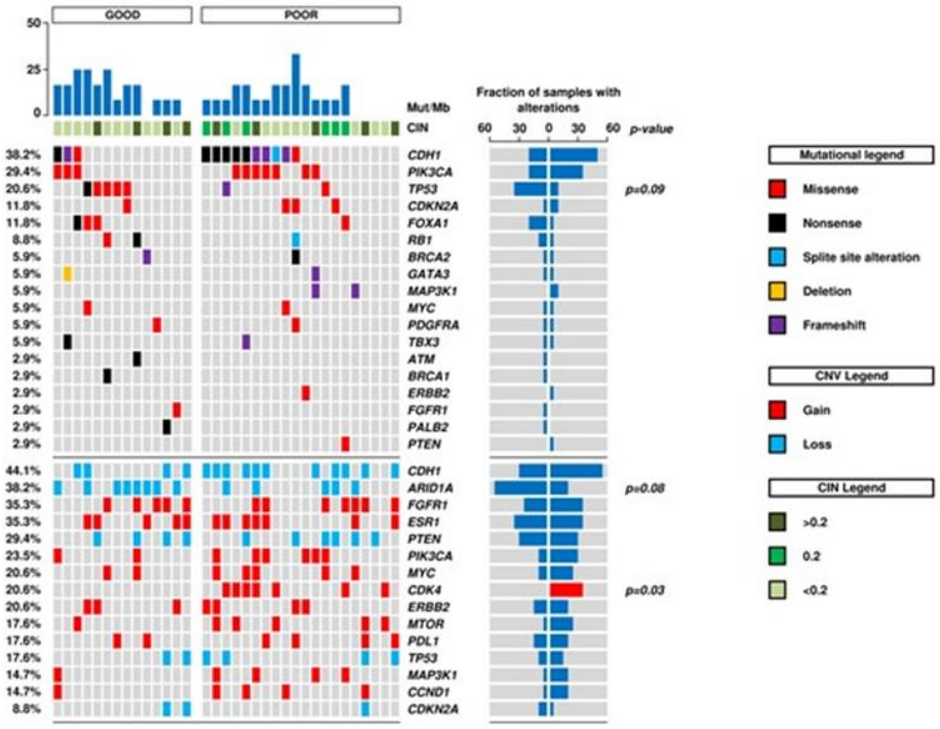
Gene	Alteration	Poor Group Patients N (%)	Good Group Patients N (%)	Total N (%)	*p-value**
AKT1	-	-	-	-	-
ARID1A	Loss	5 (25.0)	8 (57.1)	13 (38.2)	*0.08*
ATM	SM	0	1 (7.1)	1 (2.9)	-

BRCA1	SM	0	1 (7.1)	1 (2.9)	-
BRCA2	SM	1 (5.0)	1 (7.1)	2 (5.9)	-
CCND1	Gain	4 (20.0)	1 (7.1)	5 (14.7)	-
CDH1	SM	10 (50.0)	3 (21.4)	13 (38.2)	-
	Loss	11 (55.0)	4 (28.6)	15 (44.1)	-
CDK4	Gain	7 (35.0)	0	7 (20.6)	*0.03*
CDKN2A	SM	3 (15.0)	1 (7.1)	4 (11.8)	-
	Loss	1 (5.0)	2 (14.3)	3 (8.8)	-
ERBB2	SM	1 (5.0)	0	1 (2.9)	-
	Gain	4 (20.0)	3 (21.4)	7 (20.6)	-
ESR1	Gain	7 (35.0)	5 (35.7)	12 (35.3)	-
FGFR1	SM	0	1 (7.1)	1 (2.9)	-
	Gain	7 (35.0)	5 (35.7)	12 (35.3)	-
FOXA1	SM	1 (5.0)	3 (21.4)	4 (11.8)	-
	Gain	1 (5.0)	1 (7.1)	2 (5.9)	-
GATA3	SM	1 (5.0)	1 (7.1)	2 (5.9)	-
MAP3K1	SM	2 (10.0)	0	2 (5.9)	-
	Gain	4 (20.0)	1 (7.1)	5 (14.7)	-
MTOR	Gain	5 (25.0)	1 (7.1)	6 (17.6)	-
MYC	SM	1 (5.0)	1 (7.1)	2 (5.9)	-
	Gain	5 (25.0)	2 (14.3)	7 (20.6)	-
PALB2	SM	0	1 (7.1)	1 (2.9)	-
PDGFRA	SM	1 (5.0)	1 (7.1)	2 (5.9)	-
PDL1	Gain	4 (20.0)	2 (14.3)	6 (17.6)	-
PGR	-	-	-	-	-
PIK3CA	SM	7 (35.0)	3 (21.4)	10 (29.4)	-
	Gain	6 (30.0)	2 (14.3)	8 (23.5)	-
PTEN	SM	1 (5.0)	0	1 (2.9)	-
	Loss	6 (30.0)	4 (28.6)	10 (29.4)	-
RB1	SM	1 (5.0)	2 (14.3)	3 (8.8)	-
TBX3	SM	1 (5.0)	1 (7.1)	2 (5.9)	-
TP53	SM	1 (5.0)	5 (35.7)	7 (20.6)	*0.09*
	Loss	4 (20.0)	2 (14.3)	6 (17.6)	-

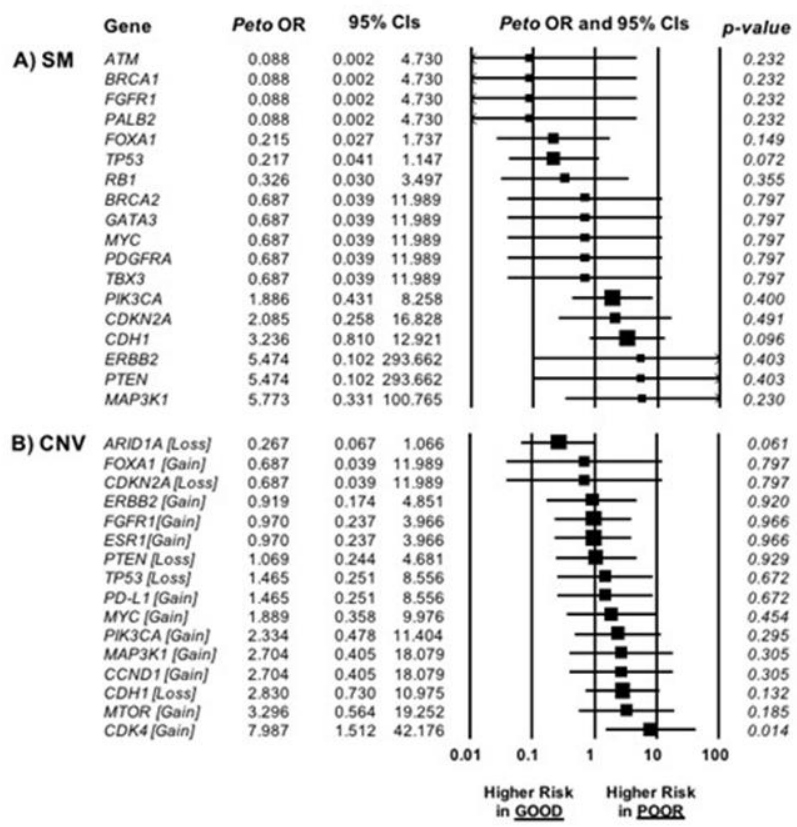
Legend Table 13. N, Number; SM, somatic mutation; p-value* according to Fisher's exact test (only *p<0.10* are reported).

Figure 9. Comparison of mutational load, chromosome integrity number, somatic mutations and copy number variation between patients with poor prognosis and patients with good prognosis [Panel A]. Odds Ratio analysis [Panel B] of somatic mutations (Panel A) and copy-number variation (Panel B) according to prognosis: an OR<1 indicates a higher chance to be associated with good prognosis; an OR>1 indicates a higher chance to be associated with poor prognosis. CIN, chromosome integrity number; CNV, copy number variation; OR, odds Ratio; CI, confidence interval; SM, somatic mutation; CNV, copy number variation.

[A]


GOOD
POOR
50
25
0
Mut/Mb
CIN
Fraction of samples with alterations
60 30 0 30 60
p-value
38.2% CDH1
29.4% PIK3CA
20.6% TP53
11.8% CDKN2A
11.8% FOXA1
8.8% RB1
5.9% BRCA2
5.9% GATA3
5.9% MAP3K1
5.9% MYC
5.9% PDGFRA
5.9% TBX3
2.9% ATM
2.9% BRCA1
2.9% ERBB2
2.9% FGFR1
2.9% PALB2
2.9% PTEN
p=0.09
44.1% CDH1
38.2% ARID1A
35.3% FGFR1
35.3% ESR1
29.4% PTEN
23.5% PIK3CA
20.6% MYC
20.6% CDK4
20.6% ERBB2
17.6% MTOR
17.6% PDL1
17.6% TP53
14.7% MAP3K1
14.7% CCND1
8.8% CDKN2A
p=0.08
p=0.03
Mutational legend
Missense
Nonsense
Splite site alteration
Deletion
Frameshift
CNV Legend
Gain
Loss
CIN Legend
>0.2
0.2
<0.2

[B]


Gene
Peto OR
95% CIs
Peto OR and 95% CIs
p-value
A) SM
ATM 0.088 0.002 4.730 0.232
BRCA1 0.088 0.002 4.730 0.232
FGFR1 0.088 0.002 4.730 0.232
PALB2 0.088 0.002 4.730 0.232
FOXA1 0.215 0.027 1.737 0.149
TP53 0.217 0.041 1.147 0.072
RB1 0.326 0.030 3.497 0.355
BRCA2 0.687 0.039 11.989 0.797
GATA3 0.687 0.039 11.989 0.797
MYC 0.687 0.039 11.989 0.797
PDGFRA 0.687 0.039 11.989 0.797
TBX3 0.687 0.039 11.989 0.797
PIK3CA 1.886 0.431 8.258 0.400
CDKN2A 2.085 0.258 16.828 0.491
CDH1 3.236 0.810 12.921 0.096
ERBB2 5.474 0.102 293.662 0.403
PTEN 5.474 0.102 293.662 0.403
MAP3K1 5.773 0.331 100.765 0.230
B) CNV
ARID1A [Loss] 0.267 0.067 1.066 0.061
FOXA1 [Gain] 0.687 0.039 11.989 0.797
CDKN2A [Loss] 0.687 0.039 11.989 0.797
ERBB2 [Gain] 0.919 0.174 4.851 0.920
FGFR1[Gain] 0.970 0.237 3.966 0.966
ESR1[Gain] 0.970 0.237 3.966 0.966
PTEN [Loss] 1.069 0.244 4.681 0.929
TP53 [Loss] 1.465 0.251 8.556 0.672
PD-L1 [Gain] 1.465 0.251 8.556 0.672
MYC [Gain] 1.889 0.358 9.976 0.454
PIK3CA [Gain] 2.334 0.478 11.404 0.295
MAP3K1 [Gain] 2.704 0.405 18.079 0.305
CCND1 [Gain] 2.704 0.405 18.079 0.305
CDH1 [Loss] 2.830 0.730 10.975 0.132
MTOR [Gain] 3.296 0.564 19.252 0.185
CDK4 [Gain] 7.987 1.512 42.176 0.014
0.01 0.1 1 10 100
Higher Risk in GOOD
Higher Risk in POOR

When grouping the molecular alterations involved in the regulation of G1/S phase cell cycle progression herein evaluated (*CDK4* and *CCND1* gain or *CDKN2A* mutation), this signature was significantly more associated with poor prognosis in the whole patients' sample (OR 6.24, 95%CI 1.59-24.5, *p=0.009*), and in the RB1 wild type population (32 patients, OR 5.33, 95% CI 1.33-21.28, *p=0.018*).
RNA was available for 20 samples (12 poor and 8 good). Ninety genes (as reported in Figure 10) showed an adjusted p-value under *0.05* and were able to discriminate all good samples to poor ones.

Figure 10. Heatmap displaying normalized expression values of the 90 differentially expressed genes between the two prognostic groups at a p-value cutoff of 0.05. Hierarchical clustering correctly separates the good and the poor samples.

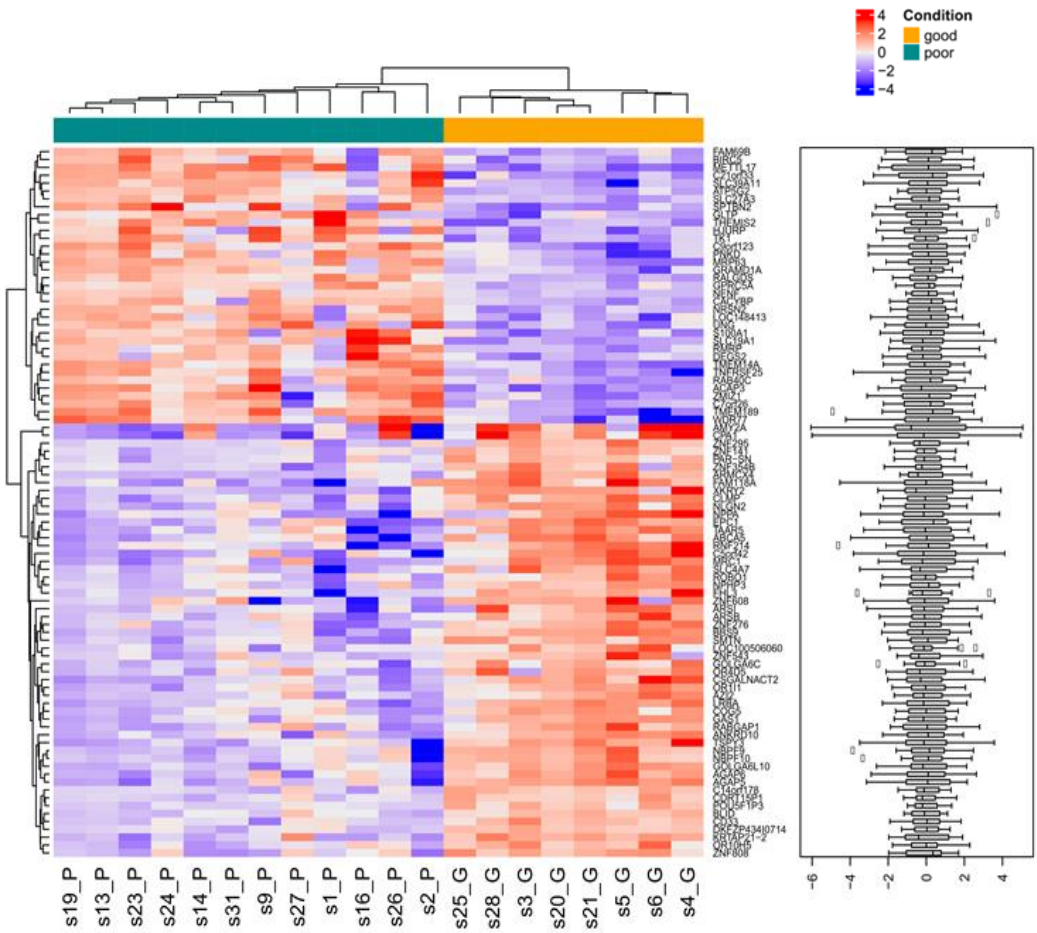


Table 14. List of differential expressed genes ordered according to decrement of adjust p-value (padj).

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
FAM118A	287.9815	-4.16484	0.71739	-5.80555	6.42E-09	6.23E-05
METTL17	11.96358	5.484037	0.95744	5.727812	1.02E-08	6.23E-05
ACAP3	54.69334	3.508042	0.641507	5.468437	4.54E-08	0.000185
ARMCX4	61.82765	-2.45036	0.502328	-4.878	1.07E-06	0.002949
ATP5G2	135.398	1.830659	0.377078	4.854862	1.2E-06	0.002949
CLMP	47.86417	-3.11953	0.65164	-4.7872	1.69E-06	0.003184
BBS9	57.56298	-2.91145	0.61287	-4.75052	2.03E-06	0.003184
SLC19A1	6.593031	5.736359	1.208823	4.745409	2.08E-06	0.003184
TNFRSF25	39.16437	3.034054	0.64686	4.690436	2.73E-06	0.003708
UNG	10.91649	3.969202	0.866673	4.579818	4.65E-06	0.005697
GAS1	486.2476	-1.83906	0.408547	-4.50148	6.75E-06	0.006523
LRBA	122.4088	-1.97506	0.439645	-4.49241	7.04E-06	0.006523
CPA1	2855.778	-4.10595	0.914598	-4.48935	7.14E-06	0.006523
GOLGA6L10	22.43287	-2.17485	0.487036	-4.46548	7.99E-06	0.006523
AGAP5	160.3393	-2.20732	0.495351	-4.45608	8.35E-06	0.006523
CSGALNACT2	59.80484	-2.45501	0.551499	-4.45153	8.53E-06	0.006523
GLTP	9.947518	3.990971	0.911693	4.377538	1.2E-05	0.008643
HJURP	20.33044	3.336638	0.766142	4.355115	1.33E-05	0.009045
XKRY2	8.999908	-4.88396	1.132215	-4.31364	1.61E-05	0.010346
NBPF9	455.8717	-1.8637	0.436331	-4.27129	1.94E-05	0.011895
MRC1	52.49322	-3.11718	0.737505	-4.22665	2.37E-05	0.013687
RNF214	79.04164	-2.61661	0.620365	-4.21786	2.47E-05	0.013687
SPTBN2	92.77476	2.947034	0.700272	4.208416	2.57E-05	0.013687
NBPF10	504.807	-1.72624	0.414547	-4.16416	3.12E-05	0.015939
TAAR5	75.10982	-2.60199	0.627422	-4.14711	3.37E-05	0.016486
C21orf33	17.91853	3.182556	0.76936	4.13663	3.52E-05	0.016593
CACYBP	243.1032	1.782038	0.43251	4.120219	3.79E-05	0.01712
C5orf42	165.0362	-2.90166	0.706717	-4.10583	4.03E-05	0.01712
ANKRD10	108.9637	-1.87962	0.458939	-4.09559	4.21E-05	0.01712
KRTAP21-2	173.7503	-2.21893	0.542046	-4.09362	4.25E-05	0.01712
TSPY3	93.7732	-2.99832	0.734317	-4.08314	4.44E-05	0.01712
WDR77	64.45564	2.922548	0.71669	4.077839	4.55E-05	0.01712
OR4D5	49.46553	-2.59136	0.636026	-4.07429	4.62E-05	0.01712
C14orf178	105.7455	-1.4312	0.351968	-4.06627	4.78E-05	0.017199
TMEM189	51.661	2.614833	0.645649	4.049929	5.12E-05	0.017877
DKFZP434I0714	237.9083	-1.28443	0.317623	-4.04388	5.26E-05	0.017877
C7orf26	49.75301	2.066796	0.512317	4.034213	5.48E-05	0.017981
ARSI	25.03895	-3.07882	0.764009	-4.02983	5.58E-05	0.017981
ZNF141	79.1625	-1.42766	0.357591	-3.99244	6.54E-05	0.020526
OR1I1	81.96834	-2.04006	0.511883	-3.98541	6.74E-05	0.020615
ZMIZ1	46.35856	2.447771	0.616872	3.968037	7.25E-05	0.021636
NPPA	27.3433	-3.99812	1.015776	-3.93602	8.28E-05	0.024145
ZNF276	25.84982	-2.63025	0.669396	-3.92928	8.52E-05	0.024254
NPHP3	15.22506	-2.30989	0.591356	-3.90609	9.38E-05	0.026096
AGAP6	166.9166	-1.93087	0.497415	-3.8818	0.000104	0.027516
TMEM14A	34.36528	2.210491	0.569804	3.879385	0.000105	0.027516
BIRC5	12.37289	3.815822	0.984161	3.877235	0.000106	0.027516
PNKD	40.58867	2.142505	0.555759	3.855101	0.000116	0.029395
FAM69B	7.964359	3.49476	0.908079	3.848519	0.000119	0.029395
SMTN	7.46489	-2.83132	0.736591	-3.84382	0.000121	0.029395
ABCA5	108.0626	-2.1638	0.563334	-3.84107	0.000123	0.029395
TK1	6.69306	4.122105	1.076128	3.830498	0.000128	0.029395
ROBO1	27.95859	-2.56238	0.669015	-3.83007	0.000128	0.029395
FHL3	90.85706	-2.0058	0.524108	-3.82708	0.00013	0.029395
NLGN2	78.29472	-2.02182	0.52942	-3.81893	0.000134	0.029831
S100A1	27.88531	2.562766	0.673243	3.806599	0.000141	0.030485
RMRP	79125.52	1.860016	0.488868	3.804738	0.000142	0.030485
THEMIS2	11.09583	3.748385	0.986994	3.797781	0.000146	0.030813
MRP63	5.867143	3.068536	0.812317	3.77751	0.000158	0.032865

SLC4A7	85.35451	-2.44587	0.649903	-3.76344	0.000168	0.034192
LOC148413	17.15092	2.726837	0.726778	3.751951	0.000175	0.034759
ZNF608	19.34841	-3.26149	0.869471	-3.75111	0.000176	0.034759
EPC1	23.07644	-2.37828	0.636215	-3.73818	0.000185	0.034844
DEGS2	15.63481	3.066862	0.820782	3.736513	0.000187	0.034844
NENF	298.75	1.252268	0.335529	3.732216	0.00019	0.034844
GRAMD1A	12.25981	1.980602	0.530794	3.731395	0.00019	0.034844
ARSB	34.92913	-2.50226	0.670665	-3.73101	0.000191	0.034844
ZNF808	186.229	-1.62569	0.437399	-3.71672	0.000202	0.035845
RAB40C	31.8133	1.902918	0.512028	3.716437	0.000202	0.035845
CD33	50.0631	-1.46189	0.394135	-3.70911	0.000208	0.036371
C9orf123	68.92144	2.434518	0.657075	3.705086	0.000211	0.036433
LOC100506060	12.6985	-2.76485	0.749924	-3.68684	0.000227	0.038418
GPRC5A	17.32872	1.883578	0.51121	3.684546	0.000229	0.038418
SLC27A3	8.187408	2.34817	0.639174	3.673756	0.000239	0.039537
NRSN2	12.68688	2.725802	0.744313	3.662171	0.00025	0.040817
CDRT15P1	177.3026	-1.30105	0.356446	-3.65007	0.000262	0.042227
RALGDS	8.410915	2.382174	0.654971	3.637069	0.000276	0.043505
SLC39A11	31.47888	2.585836	0.711835	3.632632	0.000281	0.043505
PAR-SN	191.3378	-1.31291	0.36144	-3.63243	0.000281	0.043505
ZNF354B	15.38799	-2.6369	0.726935	-3.62742	0.000286	0.043802
AMY2A	2103.127	-3.25841	0.902669	-3.60975	0.000306	0.045833
RABGAP1	56.88308	-2.5086	0.695037	-3.6093	0.000307	0.045833
BLID	104.6526	-1.26807	0.351802	-3.60451	0.000313	0.046125
GOLGA6C	71.62324	-2.00299	0.556517	-3.59915	0.000319	0.046362
ZNF543	11.90684	-2.89538	0.804947	-3.59698	0.000322	0.046362
COG5	61.98341	-1.70582	0.475018	-3.59106	0.000329	0.046407
AZI2	21.49553	-2.14747	0.59807	-3.59067	0.00033	0.046407
ZNF295	11.98757	-2.20519	0.615077	-3.58523	0.000337	0.046847
OR10H5	98.31277	-1.89321	0.528516	-3.58213	0.000341	0.046874
POU5F1P3	29.60684	-1.57231	0.441282	-3.56304	0.000367	0.049859

According to the fluorescent in situ hybridization (FISH) analysis, *CDK4* gain (Figure 11, Panel A) was detected in the 10% of patients at poor prognosis (those presenting also *CDK4* gain at NGS), while no *CDK4* gain was detected in patients at good, without a significant difference between the two groups ($p=0.5$). According to the immunohistochemistry (IHC) analysis, nuclear CDK4 overexpression (score 3+) (Figure 11, Panel B) was detected in the 20% and 7.1% of patients at poor and good prognosis, respectively ($p=0.62$). No association between *CDK4* gain at NGS and nuclear CDK4 overexpression at IHC ($p=1.0$) was observed. *CDK4* gain at FISH and nuclear CDK4 overexpression were more frequently associated with poor prognosis, despite a non-statistically significant difference, with a similar trend for CDK6 overexpression as well (Figure 11, Panel C; Figure 12). Patients with poor prognosis resulted to have a significantly higher chance to be cumulatively associated with abnormalities in CDK4/CDK6 overall expression or CNV (Figure 13).

Figure 11. Fluorescent in situ hybridization analysis for CDK4 showing a gain of the probe mapping the CDK4 gene locus (Panel A) and immunohistochemistry

analysis for nuclear CDK4 (Panel B) and CDK6 (Panel C) showing a strong immunoexpression (Score 3+).

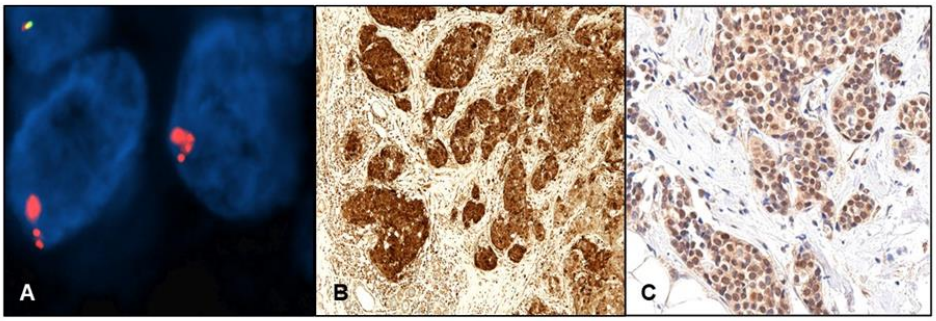


Figure 12. Odds Ratio analysis of fluorescent in situ hybridization (Panel A) and immunohistochemistry (Panel B) evaluations according to prognosis: an OR<1 indicates a higher chance to be associated with good prognosis; an OR>1 indicates a higher chance to be associated with poor prognosis. OR, odds Ratio; CI, confidence interval; FISH, fluorescent in situ hybridization; IHC, immunohistochemistry.

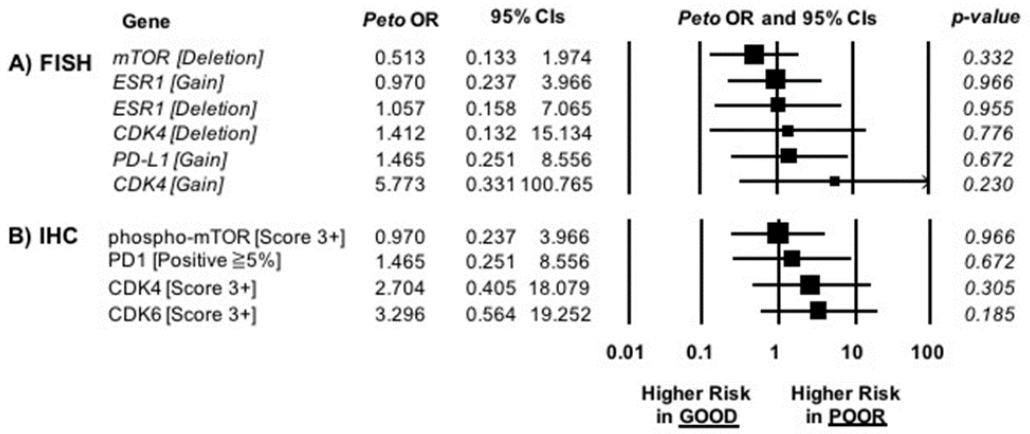
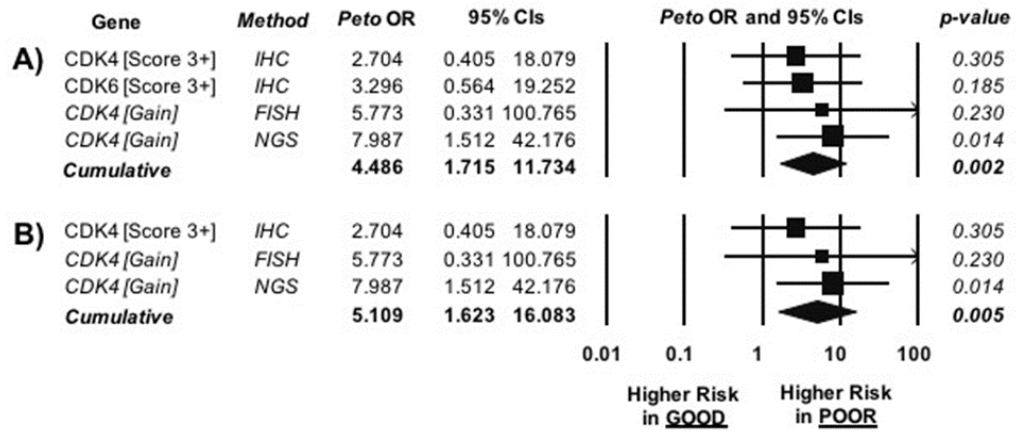


Figure 13. Cumulative Odds Ratio analysis of CDK4 (evaluated by next-generation sequencing, fluorescent in situ hybridization, and immunohistochemistry analysis) including CDK6 (evaluated by immunohistochemistry analysis) [Panel A] or excluding CDK6 (Panel B). OR, odds Ratio; CI, confidence interval; FISH, fluorescent in situ hybridization; IHC, immunohistochemistry; NGS, next-generation sequencing.



Regarding the sTIL assessment, the median value was 1% (range 0-30%), with only 6% of cases presenting a TILs percentage ≥10%. No significant association (*p=0.41*) between TILs level (considered as a categorical variable, <1% and ≥1%) and prognosis was found (OR 1.75, 95%CI 0.45-6.8, *p=0.41*). In addition, no significant association between TILs level and *CDK4* gains was identified (*p=0.99*).

DISCUSSION

Despite the improvements in the understanding of the ILC biology, the prognosis of this histotype remains controversial. The ILC still requires the identification and validation of a reliable clinical-pathological and molecular portrait, in order to stratify early-stage patients according to prognosis. In this regard, the majority of evidence regarding the potential prognostic role of proliferation derived from patients affected by IDC, and these implications are applied for ILC patients as well. This analysis suggests that the prognostic relevance of Ki67 (as well as its optimal cut-off) seems to significantly differ according to breast cancer histology. Indeed, while a threshold of 15% is indicative of high Ki67 status in IDC patients, a very low cut-off of Ki67 (4%-5%) may be able to significantly discriminate the prognosis of patients with 'pure' ILC. Therefore, these results contribute to support to not apply the prognostic Ki67 cut-off of IDC to ILC.

To date, the lack of reliable biomarkers that can identify those patients who truly benefit from cytotoxic agents represents a relevant aspect in the treatment strategy of early-stage ILC and early-stage breast cancer in general. In the context of ILC, the benefit of adjuvant chemotherapy is still unclear considering that previous studies reported no benefit of adjuvant chemotherapy in early ILC [35, 36]. Even when analyzed by multigene prognostic signatures, ILC is only rarely considered as a high-risk prognosis disease, warranting adjuvant chemotherapy [37-39].

Our propensity score analysis suggests that patients with pure luminal/HER2-negative ILC may significantly benefit from the addition of adjuvant chemotherapy to hormonotherapy in terms of OS and DDFS, particularly for large and lymph node positive tumors. A similar trend was shown for more biologic aggressive tumors, defined by high Ki67 and grading. Similar to a recent retrospective study, our results highlight that patients with high-risk ILC should not be denied adjuvant chemotherapy because of such histologic subtype [40].

In the era of cancer molecular profiling, the design and the application of risk models based on clinical parameters still provide valuable information for clinicians. Moreover, the abundance of genomic analyses does not always translate into a clinically meaningful result. Therefore, the most promising approach is likely to be represented by an integration of clinical data and genomic-proteomic

characterization. The results of this project support the strength of an integrative approach, extending from prognostic dichotomization to multi-platform genomic/transcriptomic analyses, able to unravel candidate aberrations with a biological impact on ILC oncogenesis. In this regard, the combination of clinical-pathological parameters is able to robustly stratify the prognosis of patients affected by early-stage ILC. Moreover, the external validation reinforced the discrimination performance of the prognostic model, supporting the ability of this model to accurately separate ILC patients into three risk classes according to their individual risk of recurrence. Our analysis does not represent a simple investigation of prognostic factors for ILC (as previously reported by a series of retrospective studies) [41, 42]; the novelty of this study consists in the development and validation of a prognostic tool (consisting of the combination of reliable clinical-pathological factors with different prognostic weight according to the model's score assessment) easy to adopt in the clinical practice. These results may help the clinicians to estimate the DFS and OS of early stage ILC, an histotype where several aspects (including the prognosis) represent a matter of research for the personalized medicine. Indeed, although the common prognostic factors for early-stage BC are indiscriminately applied for both lobular and ductal histology, determining that ILC patients are substantially treated in the same way as those affected by invasive carcinoma of NST, the overall prognosis of the two histotypes appears different [3, 42]. Thus, this peculiar prognostic aspect supports the hypothesis that specific molecular features might drive the ILC prognosis. In this regard, the selection of 'exceptional' responders may increase the likelihood of finding a molecular characteristic that could account for the outcome [8].

Based on this approach, we performed the molecular analysis in an explorative subset of 'worst' prognostic performers, comparing the results with those of a subset of 'best' performers. The most intriguing finding we obtained concerns the potential negative prognostic role of *CDK4* gain, detected by NGS analysis. Moreover, the *CDK4* amplification, detected by FISH analysis, and the nuclear CDK4 overexpression displayed a trend toward an association with poor prognosis. As expected in the lobular histotype, the most frequent alteration is represented by the *CDH1* mutation and loss of heterozygosity, without a different distribution in the

two prognostic groups. Similarly, mutations in *PIK3CA*, *TP53*, *CDKN2A* and *FOXA1* are not associated with prognosis. The incidence of *FOXA1* mutation and *ESR1* gain are similar to that reported by *Desmedt* et al. (11.8% *vs.* 9% and 35.3% *vs.* 25.3%, respectively) [5]. Considering that *FOXA1* plays a key role in the endocrine signaling, through the involvement in the estrogen receptor (ER)-mediated transcription, these gene alterations need to be further explored in the context of ILC, where the endocrine therapy represents the main therapeutic strategy [43]. The unsupervised clustering analysis of the transcriptome evidenced differences between the two prognostic groups in terms of gene expression level. Unlike a previous integrative study [4], our analysis showed that a series of genes, overexpressed in the poor prognosis group, may play a relevant role in the oncogenesis and treatment resistance in BC. For example, the *METTL17* interacts with both the AF1 and AF2 domains of ERα/β. In a recent study, the observation that knockdown of *METTL17* reduces BC cell growth suggests that *METTL17* regulates the cancer cell growth possibly through modulation of ERα function as well as the expression of ERα/β target genes [44].

With regard to the presence of sTIL, our analysis showed that the majority of ILCs are characterized by low levels of sTIL, without a statistically significant distribution between the two prognostic cohorts. These results are consistent with a recent analysis reporting that the percentage of sTIL in ILC was lower compared to that in invasive carcinoma of NST and that the sTIL level did not represent an independent prognostically variable [45].

CONCLUSION

In conclusion, our multi-step analyses suggested that: 1) The already known Ki67 value of 14% represents a good prognostic cut-off for patients with IDC, while a Ki67 value of 4% or 5% discriminates the prognosis of patients affected by ILC; 2) Patients affected by luminal ILC could derive significant survival benefit from the addition of chemotherapy to hormonotherapy, especially for high-risk patients; 3) A risk stratification model able to accurately separate early-stage ILC patients' prognosis into different risk classes, according to clinical-pathological variables, allowed to investigate potential biomarkers of prognosis with targeted NGS in an explorative cohort. In particular, *CDK4* gain is suggested for future validation as a prognostic biomarkers and potential therapeutic opportunity.

The introduction and validation of a personalized approach in the context of ILC might allow the clinicians to provide the best available therapy for every individual patient to potentiate the expected clinical benefit and reduce the human cost.

REFERENCES

1. Arpino G, Bardou VJ, Clark GM, Elledge RM. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. Breast Cancer Res 2004; 6: R149-156.
2. McCart Reed AE, Kutasovic JR, Lakhani SR, Simpson PT. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. Breast Cancer Res 2015; 17: 12.
3. Pestalozzi BC, Zahrieh D, Mallon E et al. Distinct clinical and prognostic features of infiltrating lobular carcinoma of the breast: combined results of 15 International Breast Cancer Study Group clinical trials. J Clin Oncol 2008; 26: 3006-3014.
4. Ciriello G, Gatza ML, Beck AH et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell 2015; 163: 506-519.
5. Desmedt C, Zoppoli G, Gundem G et al. Genomic Characterization of Primary Invasive Lobular Breast Cancer. J Clin Oncol 2016; 34: 1872-1881.
6. Michaut M, Chin SF, Majewski I et al. Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. Sci Rep 2016; 6: 18517.
7. Burstein HJ, Curigliano G, Loibl S et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. Ann Oncol 2019; 30: 1541-1557.
8. Kummar S, Williams PM, Lih CJ et al. Application of molecular profiling in clinical trials for advanced metastatic cancers. J Natl Cancer Inst 2015; 107.
9. Shuster JJ. Median follow-up in clinical trials. J Clin Oncol 1991; 9: 191-192.
10. Horton T LB. On the exact distribution of maximally selected rank statistics. . Comput Statist Data Anal. 2003; 121-137.
11. Hess KR. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. Stat Med 1995; 14: 1707-1723.
12. Harrell FE, Jr., Lee KL, Califf RM et al. Regression modelling strategies for improved prognostic prediction. Stat Med 1984; 3: 143-152.
13. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. Stat Med 1992; 11: 2093-2109.
14. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008; 26: 1364-1370.
15. Lazar AA, Cole BF, Bonetti M, Gelber RD. Evaluation of treatment-effect heterogeneity using biomarkers measured on a continuous scale: subpopulation treatment effect pattern plot. J Clin Oncol 2010; 28: 4539-4544.
16. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70: 41-55.
17. Armstrong K. Methods in comparative effectiveness research. J Clin Oncol 2012; 30: 4208-4214.
18. Bria E, Di Modugno F, Sperduti I et al. Prognostic impact of alternative splicing-derived hMENA isoforms in resected, node-negative, non-small-cell lung cancer. Oncotarget 2014; 5: 11054-11063.
19. Alonzo TA. Standards for reporting prognostic tumor marker studies. J Clin Oncol 2005; 23: 9053-9054.
20. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009; 101: 1446-1452.

21. Bria E, De Manzoni G, Beghelli S et al. A clinical-biological risk stratification model for resected gastric cancer: prognostic impact of Her2, Fhit, and APC expression status. Ann Oncol 2013; 24: 693-701.
22. Carbognin L, Sperduti I, Ciccarese M et al. Prognostic model for advanced breast carcinoma with luminal subtype and impact of hormonal maintenance: Implications for post-progression and conditional survival. Breast 2016; 29: 24-30.
23. Simbolo M, Gottardi M, Corbo V et al. DNA qualification workflow for next generation sequencing of histopathological samples. PLoS One 2013; 8: e62692.
24. Nik-Zainal S, Davies H, Staaf J et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016; 534: 47-54.
25. Cingolani P, Patel VM, Coon M et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet 2012; 3: 35.
26. McLaren W, Pritchard B, Rios D et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010; 26: 2069-2070.
27. Robinson JT, Thorvaldsdottir H, Winckler W et al. Integrative genomics viewer. Nat Biotechnol 2011; 29: 24-26.
28. Boeva V, Popova T, Lienard M et al. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. Bioinformatics 2014; 30: 3443-3450.
29. Campesato LF, Barroso-Sousa R, Jimenez L et al. Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. Oncotarget 2015; 6: 34221-34227.
30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.
31. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 2014; 32: 896-902.
32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26: 139-140.
33. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016; 32: 2847-2849.
34. Salgado R, Denkert C, Demaria S et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol 2015; 26: 259-271.
35. Marmor S, Hui JYC, Huang JL et al. Relative effectiveness of adjuvant chemotherapy for invasive lobular compared with invasive ductal carcinoma of the breast. Cancer 2017; 123: 3015-3021.
36. Truin W, Voogd AC, Vreugdenhil G et al. Effect of adjuvant chemotherapy in postmenopausal patients with invasive ductal versus lobular breast cancer. Ann Oncol 2012; 23: 2859-2865.
37. Kizy S, Huang JL, Marmor S et al. Impact of the 21-gene recurrence score on outcome in patients with invasive lobular carcinoma of the breast. Breast Cancer Res Treat 2017; 165: 757-763.
38. Sparano JA, Gray RJ, Makower DF et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med 2018; 379: 111-121.
39. Tsai ML, Lillemoe TJ, Finkelstein MJ et al. Utility of Oncotype DX Risk Assessment in Patients With Invasive Lobular Carcinoma. Clin Breast Cancer 2016; 16: 45-50.

40. de Nonneville A, Jauffret C, Goncalves A et al. Adjuvant chemotherapy in lobular carcinoma of the breast: a clinicopathological score identifies high-risk patient with survival benefit. Breast Cancer Res Treat 2019; 175: 379-387.
41. Adachi Y, Ishiguro J, Kotani H et al. Comparison of clinical outcomes between luminal invasive ductal carcinoma and luminal invasive lobular carcinoma. BMC Cancer 2016; 16: 248.
42. Chen Z, Yang J, Li S et al. Invasive lobular carcinoma of the breast: A special histological type compared with invasive ductal carcinoma. PLoS One 2017; 12: e0182397.
43. Ross-Innes CS, Stark R, Teschendorff AE et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 2012; 481: 389-393.
44. Du P, Yuan B, Cao J et al. Methyltransferase-like 17 physically and functionally interacts with estrogen receptors. IUBMB Life 2015; 67: 861-868.
45. Desmedt C, Salgado R, Fornili M et al. Immune Infiltration in Invasive Lobular Breast Cancer. J Natl Cancer Inst 2018.