

## Mitotic read-out genes confer poor outcome in luminal A breast cancer tumors

Javier Pérez-Peña<sup>1,\*</sup>, Ana Alcaraz-Sanabria<sup>1,\*</sup>, Cristina Nieto-Jiménez<sup>1,\*</sup>, Raquel Páez<sup>1</sup>, Verónica Corrales-Sánchez<sup>1</sup>, Leticia Serrano-Oviedo<sup>1</sup>, Vikram B. Wali<sup>2</sup>, Gauri A. Patwardhan<sup>2</sup>, Eitan Amir<sup>3</sup>, Balázs Győrffy<sup>4,5</sup>, Atanasio Pandiella<sup>6</sup>, Alberto Ocaña<sup>1</sup>

<sup>1</sup>Translational Research Unit, Albacete University Hospital and Centro Regional de Investigaciones Biomedicas (CRIB), Castilla La Mancha University, Albacete, Spain

<sup>2</sup>Yale Cancer Center, Yale University, New Haven, USA

<sup>3</sup>Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, University of Toronto, Toronto, Canada

<sup>4</sup>Semmelweis University, 2nd Department of Pediatrics, Budapest, Hungary

<sup>5</sup>MTA TTK Lendület Cancer Biomarker Research Group, Budapest, Hungary

<sup>6</sup>Cancer Research Center and CIBERONC, CSIC-University of Salamanca, Salamanca, Spain

\*These authors have contributed equally to this work

Correspondence to: Alberto Ocaña, email: albertoo@sescam.jccm.es

Keywords: breast cancer, luminal A, mitotic kinases, clinical outcome

Received: October 19, 2016

Accepted: January 16, 2017

Published: February 21, 2017

### ABSTRACT

Luminal breast tumors have been classified into A and B subgroups, with the luminal A being associated with a more favorable clinical outcome. Unfortunately, luminal A tumors do not have a universally good prognosis. We used transcriptomic analyses using public datasets to evaluate the differential expression between normal breast tissue and breast cancer, focusing on upregulated genes included in cell cycle function. Association of selected genes with relapse free survival (RFS) and overall survival (OS) was performed using the KM Plotter Online Tool (<http://www.kmplot.com>). Seventy-seven genes were differentially expressed between normal and malignant breast tissue. Only five genes were associated with poor RFS and OS. The mitosis-related genes GTSE1, CDCA3, FAM83D and SMC4 were associated with poor outcome specifically in Luminal A tumors. The combination of FAM83D and CDCA3 for RFS and GTSE1 alone for OS showed the better prediction for clinical outcome. CDCA3 was amplified in 3.4% of the tumors, and FAM83D and SMC4 in 2.3% and 2.2%, respectively. In conclusion, we describe a set of genes that predict detrimental outcome in Luminal A tumors. These genes may have utility for stratification in trials of antimitotic agents or cytotoxic chemotherapy, or as candidates for direct target inhibition.

### INTRODUCTION

Breast cancer has been classified in different subtypes by gene expression analyses, and each of these subtypes is associated with different clinical outcomes [1, 2]. The luminal subtype includes those tumors that express the estrogen receptor, and therefore can be targeted with hormonal therapy [1, 3]. This subgroup has been classified as luminal A and B, and similarly these two subtypes are associated with different clinical outcomes [1, 4]. Luminal B tumors are linked with a more aggressive phenotype while luminal A shows a more benign behavior [4]. Luminal B tumors are characterized by an increased

in proliferation, which is evidenced by elevated levels of Ki67 [5]. In addition, Luminal B tumors respond better to cytotoxic chemotherapy than the luminal A subtype.

It is known that the classification of breast cancer can stratify risk and prognosis; however not all tumors within a subgroup have similar clinical behavior. For instance, a subset of luminal A cancers are associated with poor outcome. Therefore the identification of these tumors could help to optimize therapy and to explore novel therapeutic strategies.

One of the main differences between the two luminal subgroups is the capacity of tumors to proliferate. However, measures of proliferation such as Ki67 are

imperfect and some tumors that depend of mitosis are not identified. Therefore, the recognition of luminal A tumors that are mitotically more active and associated with worse outcome could help clinicians to customize their therapeutic decisions.

We hypothesize that by evaluating dysregulated genes involved in cell cycle and that are overexpressed in other breast cancer subtypes, we can identify a set of genes that would permit us to select those luminal A tumors associated with worse outcome. The identification of this subgroup would have important clinical implications as it could help determine which patients will respond poorly to endocrine therapy, opening the possibility to explore other therapeutic options in this group.

## RESULTS

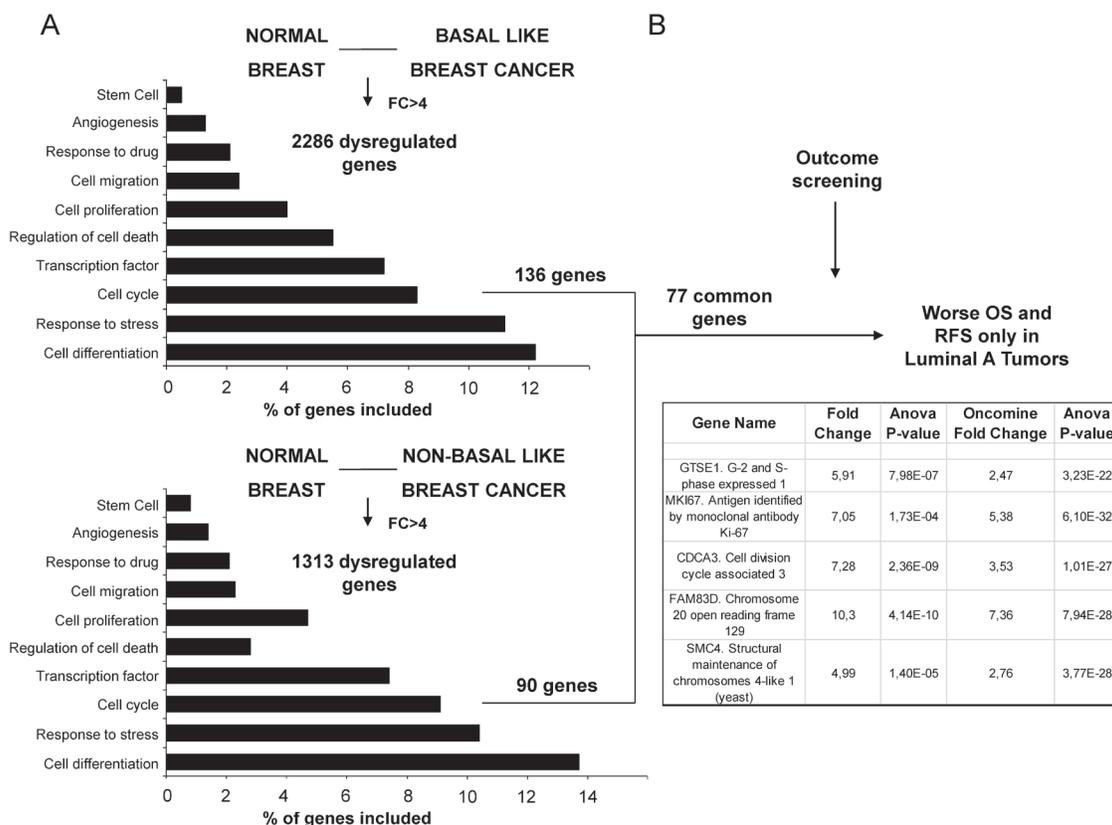
### Transcriptomic analysis identify upregulated genes linked with cell cycle

By comparing gene expression data from normal breast tissue and breast cancers, using a minimum fold change of 4, we identified 136 and 90 dysregulated genes

included in the cell cycle function, in basal and non-basal breast cancers, respectively. Functional genomics using DAVID bioinformatic resources 6.7 identified several functions among the gene candidates (Figure 1A). A total number of 77 genes associated with cell cycle were shared among both basal and non-basal cancers (Figure 1A).

### Mitotic-related genes GTSE, CDCA3, FAM83D and SMC4 are associated with poor outcome

Using the online tool KM plotter (<http://www.kmplot.com>) [6] we selected genes that were associated with detrimental outcome specifically in luminal A tumors. The definition of breast cancer subtypes is described in material and methods. Among the 77 genes identified only 5 were associated with poor relapse free survival (RFS) and overall survival (OS) in the luminal A subtype (Figure 1B). One of the identified genes was MKi67, the gene that codes for Ki67. Therefore, we did not consider it for further outcome analyses as it is the gene used in the clinical setting to discriminate luminal A from luminal B subgroups. Of note, the predictive capacity for this gene was lower than the others selected, mainly for OS



**Figure 1: Identification of genes associated with detrimental outcome in Luminal A tumors.** A. Transcriptomic expression and pathway analyses among normal breast and basal-like and non-basal-like breast cancers, with the identification of dysregulated genes with more than  $\geq 4$  fold change included in the cell cycle gene ontology categories. B. Outcome screening for detrimental relapse free survival and overall survival. Confirmation of expression values between normal breast and breast cancer using data contained at Oncomine.

(HR: 1.4, CI: 1.17-1.67; log-rank  $p=0.00023$  and HR: 1.47, CI: 1.01-2.16; log-rank  $p=0.046$ , for RFS and OS, respectively).

The four selected genes were GTSE1, CDCA3, FAM83D and SMC4. Each gene was associated with detrimental RFS and OS in luminal A tumors as shown in Figure 2 and 3, respectively (additional data is shown in Supplementary Table 1 and 2, respectively). The upregulated expression of these genes in breast cancer was confirmed using data contained at Oncomine (Figure 1B).

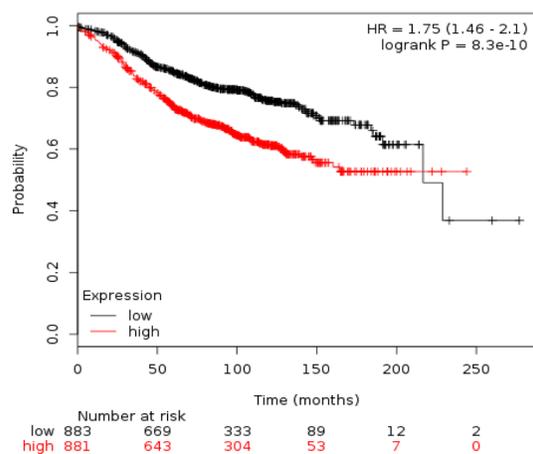
### Combined analyses of GTSE1, CDCA3, FAM83D and SMC4 predicts poor RFS and OS in Luminal A tumors

When combined, the expression of GTSE1, CDCA3, FAM83D and SMC4 was significantly associated with worse RFS (HR 2.14, CI: 1.64-2.8; log rank  $p=1.2e-08$ ) (Figure 4). Additionally, the combination of CDCA3 and

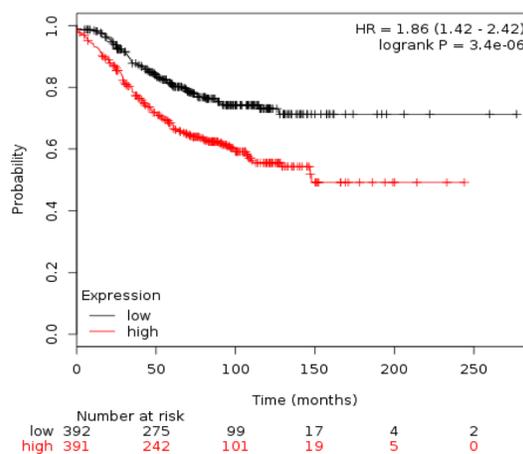
FAM83D showed the highest effect on RFS (HR 2.34, CI: 1.78-3.08; log rank  $p=3.60e-10$ ) (Table 1A). No association between these genes and RFS was observed for luminal B tumors (Supplementary Table 1). Similarly, a detrimental effect was observed for OS with the combined analyses of the four genes (HR 2.3; CI: 1.27-4.18; log rank  $p=0.0049$ ) (Figure 5), but GTSE1 alone predicted the greatest magnitude of effect (HR 3.23; CI: 2.11-4.94; log rank  $p=1.2e-08$ ) (Figure 3, and Supplementary Table 2). Of note, neither combination among genes was better than FAM83D and CDCA3 for RFS, and GTSE1 for OS (Table 1A and 1B, and Supplementary Table 2, respectively).

### Association with outcome by nodal status and treatment with chemotherapy

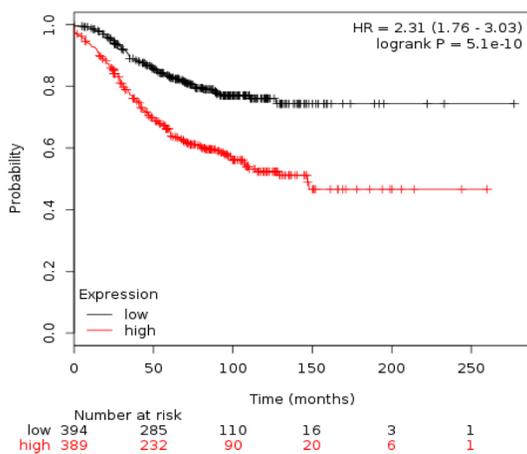
Next we evaluated the association of these genes alone or in combination with outcome by nodal status and chemotherapy treatment. GTSE1 and the combination



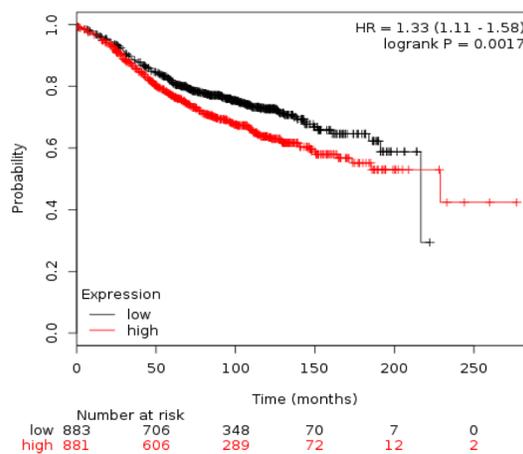
GTSE1. G-2 and S-phase expressed 1



CDCA3. Cell division cycle associated 3

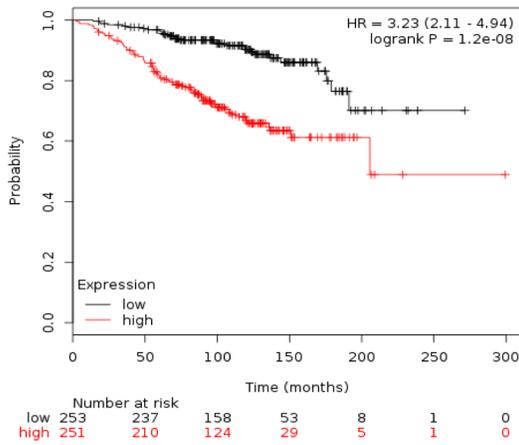


FAM83D. Chromosome 20 open reading frame 129

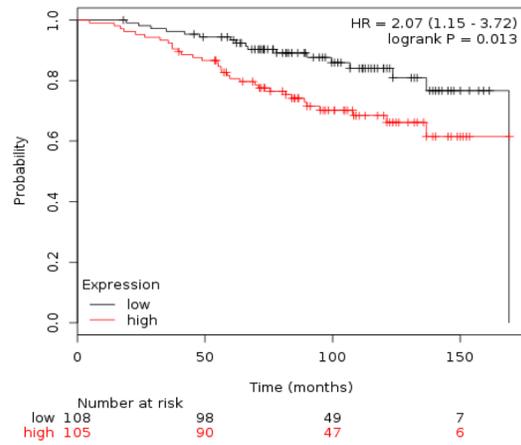


SMC4. Structural maintenance of chromosomes 4-like 1 (yeast)

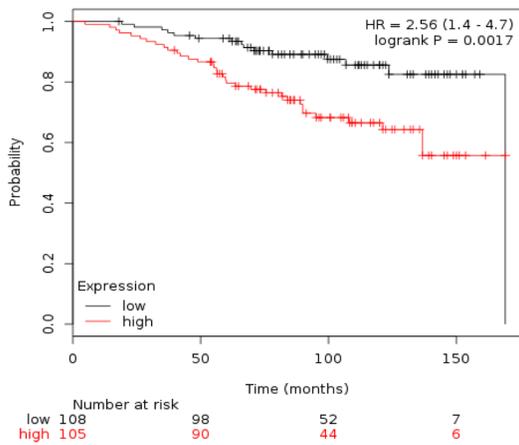
Figure 2: Association of GTSE1, CDCA3, FAM83D and SMC4 individually with relapse free survival in Luminal A tumors using KM Plotter online tool, as described in material and methods.



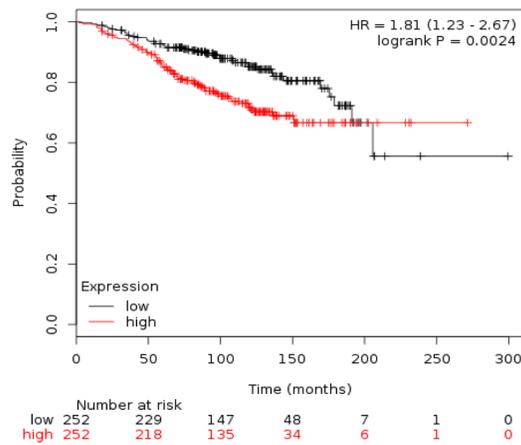
GTSE1. G-2 and S-phase expressed 1



CDCA3. Cell division cycle associated 3

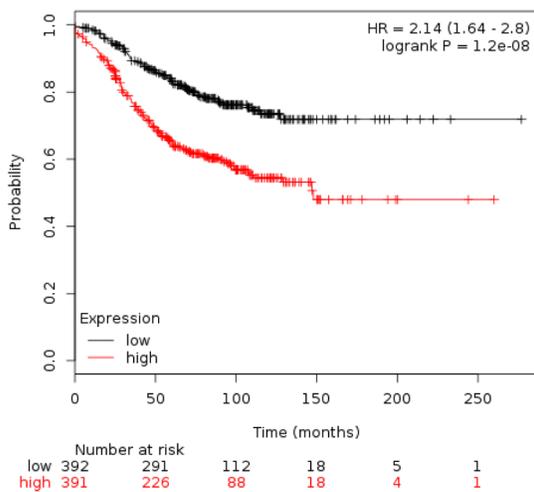


FAM83D. Chromosome 20 open reading frame 129

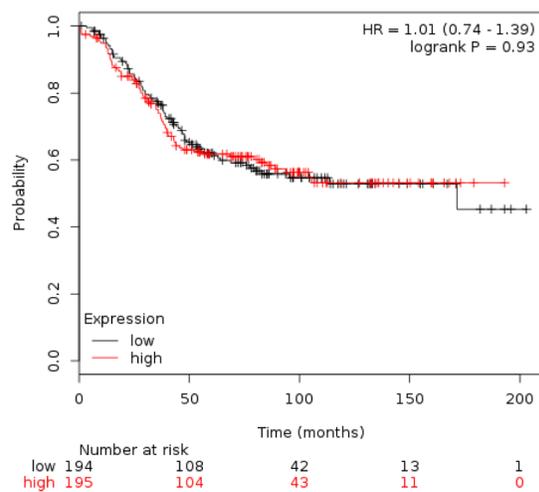


SMC4. Structural maintenance of chromosomes 4-like 1 (yeast)

Figure 3: Association of GTSE, CDCA3, FAM83D and SMC4 individually with overall survival in Luminal A tumors, using KM Plotter online tool, as described in material and methods.



Luminal A



Luminal B

Figure 4: Association of the combined analyses of GTSE, CDCA3, FAM83D and SMC4 with relapse free survival in Luminal A. and B. tumors using KM Plotter online tool, as described in material and methods.

**Table 1: A. Association with Relapse Free Survival of gene combinations in Luminal A tumors. B. Association with Overall Survival of gene combinations in Luminal A tumors**

**A**

Relapse Free Survival. Luminal A		
Gene Symbols	Hazard Ratio	P-value
GTSE1 + CDCA3	1.82	7.40E-06
GTSE1 + FAM83D	2.09	3.80E-08
GTSE1 + SMC4	1.41	1.40E-04
CDCA3 + FAM83D	2.34	3.60E-10
CDCA3 + SMC4	1.7	6.20E-05
FAM83D + SMC4	1.85	3.40E-06
CDCA3 + FAM83D + GTSE1	2.03	1.40E-07
CDCA3 + FAM83D + SMC4	1.97	3.60E-07
FAM83D + GTSE1 + SMC4	1.91	9.90E-07

**B**

Overall Survival. Luminal A		
Gene Symbols	Hazard Ratio	P-value
GTSE1 + CDCA3	2.02	0.017
GTSE1 + FAM83D	2.48	2.40E-03
GTSE1 + SMC4	2.13	1.30E-04
CDCA3 + FAM83D	2.22	7.20E-03
CDCA3 + SMC4	1.99	0.019
FAM83D + SMC4	2.17	9.10E-03
CDCA3 + FAM83D + GTSE1	2.57	1.60E-03
CDCA3 + FAM83D + SMC4	2.26	6.00E-03
FAM83D + GTSE1 + SMC4	2.52	2.00E-03

KM Plotter online tool was used as described in material and methods

of GTSE1, SMC4, FAM83D and CDCA3 predicted detrimental outcome for RFS in patients with both axillary positive and negative lymph nodes (Supplementary Table 3). Expression of GTSE1 was associated with poor OS in both subgroups. Similarly, GTSE1 predicted for detrimental RFS and OS in patients treated with or without chemotherapy, and the combination of the four genes predicted for RFS (Supplementary Table 3).

### Molecular alterations or copy number gains in the evaluated genes and potential druggable targets

Supplementary Table 4 describes the function of these genes. Finally, we evaluated if the association of these genes with outcome could be related to molecular alterations such as mutations or copy number

modifications. CDCA3 was amplified in 3.4% of the tumors, and FAM83D and SMC4 in 2.3% and 2.2%, respectively (Supplementary Table 5). Finally using the Drug Gene Interaction Database (DGIdb) (<http://dgidb.genome.wustl.edu/>) we explored compounds that could potentially interact with the identified genes. Only the compound hydrochlorothiazide showed potential for interaction with CDCA3. No compound was found for potential interaction with GTSE1, FAM83D or SMC4.

### DISCUSSION

We have discovered a set of genes that identify luminal A patients with worse outcome. It is known that most patients that harbor luminal A tumors are treated with endocrine therapy. However, a subset of these tumors does not respond adequately to this treatment and have

poor prognosis. In this context, the identification of a subgroup of patients within the luminal A subtype with a poor prognosis could help to stratify patients and consider alternative therapeutic strategies for them.

In our study by using transcriptomic analyses we identified genes that are overexpressed in breast cancer and linked with cell division. Among the genes commonly shared in the analyses, only four were associated with poor RFS and OS in luminal A tumors. Of note, the analyses of GTSE1 for OS and FAM83D and CDCA3 for RFS showed the worse outcome, an observation seen only in luminal A tumors.

CDCA3, FAM83D and SMC4 genes were found to be amplified in a very small percentage of tumors. The reduced presence of molecular alterations or copy modifications at a genomic level, lead us to consider these markers as an indirect measure of tumors with high proliferation capacity. It can be considered that these genes are not oncogenic drivers but may be indirect indicators of tumors that are more dependent of cell division and mitosis. In this context, tumors with these markers could be more sensitive to agents that target cell cycle or even cytotoxic chemotherapy. Therefore, these genes may have utility in identifying patients for such treatments.

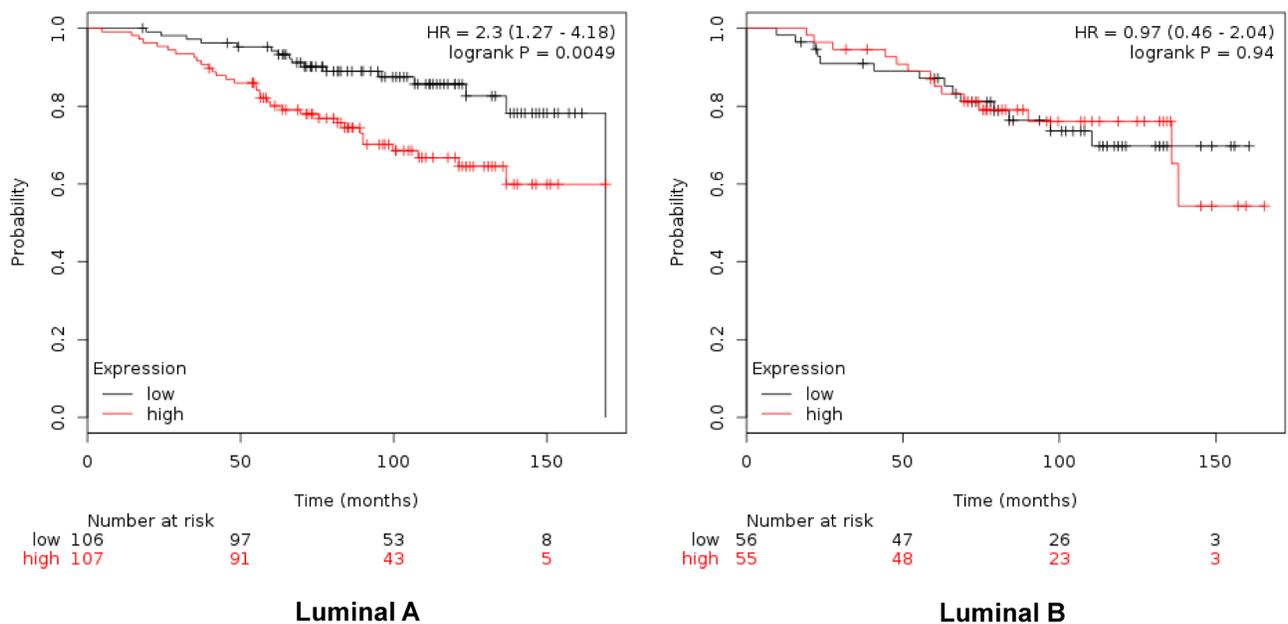
CDCA3 has been associated with both cancer risk [7] and poor prognosis in certain tumors [8]. High levels of FAM83D have been associated with poor outcome in several tumors and an increase in proliferation *in vitro* models [9, 10]. SMC4 belongs to a family of genes linked with poor outcome in prostate cancer [11], and finally, GTSE1 was described as associated with worse prognosis in uterine leiomyosarcoma [12].

It should be mentioned that one of the identified genes that was differentially expressed was MKi67, the gene that encodes for Ki67. As this gene is currently used to select between luminal A and B tumors, therefore it was not included in our evaluation. However, the magnitude of association between the expression of MKi67 and poor outcome was lower than the selected genes.

In addition it should be noted that the genes identified are not druggable targets as can be seen from our analyses, with the exception of hydrochlorothiazide that could have a potential interaction with CDCA3.

A limitation of our study is a potential heterogeneity in the treatment the patients received. Although it can be expected that most estrogen-positive patients were treated with endocrine therapy, this cannot be confirmed. Also, the results are based on univariable analysis as there was insufficient information on patient and tumor characteristics to inform a multivariable model. This, in addition to survival data which were available for only a small number of patients, suggests that results should be considered hypotheses generating and should be validated in a large independent cohort. Our analysis focused on proliferation. It is known that other biological functions beyond mitosis can differentiate between luminal A and B tumors. Finally, it should be mentioned that information contained at Oncomine does not distinguish among breast cancer subtypes.

In conclusion, we describe a set of genes that are overexpressed in luminal A tumors that predict for detrimental outcome. These biomarkers could help to stratify therapies including treatment with antimetabolic agents or cytotoxic chemotherapy.



**Figure 5: Association of the combined analyses of GTSE, CDCA3, FAM83D and SMC4 with overall survival in Luminal A. and B. tumors using KM Plotter online tool, as described in material and methods.**

## MATERIALS AND METHODS

### Transcriptomic analysis and identification of upregulated genes

mRNA level data from normal breast tissue and basal-like and non-basal like breast tumors were extracted from a public dataset (GEO DataSet accession number: GDS2250)[13, 14]. Affymetrix CEL files were downloaded and analyzed with Affymetrix Transcriptome Analysis Console 3.0. Only genes with minimum 4-fold differential expression values between control and other groups were selected. The list of genes was analyzed using gene set enrichment analyses using DAVID Bioinformatics Resources 6.7 in order to identify functions of these genes (<https://david.ncifcrf.gov/>). We used an adjusted p-value <0.05 to select the enriched gene-sets. The differentially expressed genes were independently confirmed using data contained at Oncomine ([www.oncomine.org](http://www.oncomine.org)) (TCGA Breast). For this analysis, due to the absence of breast cancer subtypes in this dataset we compared normal breast with breast cancer.

### Outcome analyses

The KM Plotter Online Tool (<http://www.kmplot.com>)[6] was used to evaluate the relationship between the presence of different genes and patient clinical outcome in different breast cancer subtypes.

This publicly available database allows to investigate overall survival (OS) and relapse-free survival (RFS) in luminal A, luminal B and basal-like breast cancers.

### Definitions of breast cancer subtypes

Breast cancer subtypes in the KM Plotter Online tool are defined as follow: Triple negative: ESR1-/HER2-. Luminal A: ESR1+/HER2-/MKI67 low. Luminal B: ESR1+/HER2-/MKI67 high and ESR1+/HER2+ -.

### Evaluation of molecular alterations

We used data contained at cBioportal ([www.cbioportal.org](http://www.cbioportal.org)) (TCGA dataset) to explore the role of mutations, deletions or amplifications in the identified genes [15].

### Evaluation of gene-drug interactions

For the evaluation of compounds that could potentially interact with the identified genes we used the Drug Gene Interaction Database (DGIdb) (<http://dgidb.genome.wustl.edu/>).

## ACKNOWLEDGMENTS

Instituto de Salud Carlos III (PI16/01121), ACEPAIN; Diputación de Albacete and CRIS Cancer Foundation (to AO). BAE (Beca Ampliación de Estudios) to AO for his stay at Yale University. Ministry of Economy and Competitiveness of Spain (BFU2012-39151 and BFU2015-71371-R), the Instituto de Salud Carlos III through the Spanish Cancer Centers Network Program (RD12/0036/0003) and the CIBERONC, the scientific foundation of the AECC and the CRIS Foundation (to AP). The work carried out in the Spanish laboratories receives support from the European Community through the regional development funding program (FEDER).

## CONFLICTS OF INTEREST

No conflicts of interest to declare for any author of this manuscript.

### Author contributions

Javier Pérez-Peña, Ana Alcaraz-Sanabria, Cristina Nieto-Jiménez, Raquel Páez, Verónica Corrales-Sánchez, Leticia Serrano-Oviedo, Balázs Györfy: collection and analyses of data.

Vikram B. Wali, Gauri Patwardhan, Eitan Amir, Atanasio Pandiella, Balázs Györfy and Alberto Ocaña: manuscript writing.

Alberto Ocana: study design and idea.

All authors approved the final version of the manuscript.

## REFERENCES

1. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98:10869-10874.
2. Ocana A, Pandiella A. Targeting oncogenic vulnerabilities in triple negative breast cancer: biological bases and ongoing clinical studies. *Oncotarget*. 2017. doi: 10.18632/oncotarget.14731. [Epub ahead of print].
3. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406:747-752.
4. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003; 100:8418-8423.

5. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ, Nielsen TO. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst.* 2009; 101:736-750.
6. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat.* 2010; 123:725-731.
7. Uchida F, Uzawa K, Kasamatsu A, Denkert C, Budczies J, Li Q, Szallasi Z. Overexpression of cell cycle regulator CDCA3 promotes oral cancer progression by enhancing cell proliferation with prevention of G1 phase arrest. *BMC Cancer.* 2012; 12:321.
8. Itzel T, Scholz P, Maass T, Krupp M, Marquardt JU, Strand S, Becker D, Staib F, Binder H, Roessler S, Wang XW, Thorgeirsson S, Müller M, et al. Translating bioinformatics in oncology: guilt-by-profiling analysis and identification of KIF18B and CDCA3 as novel driver genes in carcinogenesis. *Bioinformatics.* 2015; 31:216-224.
9. Walian PJ, Hang B, Mao JH. Prognostic significance of FAM83D gene expression across human cancer types. *Oncotarget.* 2016; 7:3332-3340. doi: 10.18632/oncotarget.6620.
10. Wang Z, Liu Y, Zhang P, Zhang W, Wang W, Curr K, Wei G, Mao JH. FAM83D promotes cell proliferation and motility by downregulating tumor suppressor gene FBXW7. *Oncotarget.* 2013; 4:2476-2486. doi: 10.18632/oncotarget.1581.
11. Zhao SG, Evans JR, Kothari V, Sun G, Larm A, Mondine V, Schaeffer EM, Ross AE, Klein EA, Den RB, Dicker AP, Karnes RJ, Erho N, et al. The Landscape of Prognostic Outlier Genes in High-Risk Prostate Cancer. *Clin Cancer Res.* 2016; 22:1777-1786.
12. Barlin JN, Zhou QC, Leitao MM, Bisogna M, Olvera N, Shih KK, Jacobsen A, Schultz N, Tap WD, Hensley ML, Schwartz GK, Boyd J, Qin LX, et al. Molecular subtypes of uterine leiomyosarcoma and correlation with clinical outcome. *Neoplasia.* 2015; 17:183-189.
13. Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell.* 2006; 9:121-132.
14. Alimonti A, Carracedo A, Clohessy JG, Trotman LC, Nardella C, Egia A, Salmena L, Sampieri K, Haveman WJ, Brogi E, Richardson AL, Zhang J, Pandolfi PP, et al. Subtle variations in Pten dose determine cancer susceptibility. *Nat Genet.* 2010; 42:454-458.
15. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013; 6:p11. doi: scisignal. 2004088.