Risk score based on expression of AP2-α, CHI3L1 and PBEF1 is a better predictor of prognosis in breast cancer

Umaira [1]Tabassum, Irene Rosita Pia [2]Patric, Geetashree [1*]Mukherjee and Kumaravel [2*]Somasundaram

Department of Microbiology and Cell Biology[2], Indian Institute of Science, Bangalore 12, Department of [1]Pathology, Kidwai Memorial Institute of Oncology, Bangalore 68

Abstract- Modulation of AP-2α, AP-2γ, CHI3L1 and PBEF1 has been implicated in the pathogenesis of many human neoplasms. In order to get an insight into the role of these genes in breast cancer pathogenesis, we have studied the extent of AP-2α, AP-2γ, CHI3L1 and PBEF1 expression and evaluated their prognostic significance by calculating risk score among Indian sporadic breast cancer patients. Immunohistochemistry was carried out to study the expression level of four genes AP-2α, AP-2γ, CHI3L1 and PBEF1. Statistical methods were used to calculate risk score which was tested for prognostic significance. While AP-2α (nuclear), CHI3L1, PBEF1, ErbB2, Stage, distant metastasis and nodal involvement correlated with patient survival in univariate analysis, multivariate analysis revealed that AP-2α, CHI3L1, PBEF1, ErbB2, distant metastasis and lymph node involvement were independent predictors of survival. Of the markers of our interest, AP-2α (nuclear) was found to be protective, whereas CHI3L1 and PBEF1 were found to be risky. A risk score calculated by combining the prognostic value of these three markers was found to be a significant predictor of survival by univariate analysis (HR = 2.532; 95% CI = 1.823-3.516; P = $2.912x10^{-8}$). The risk score segregated the patients into high and low risk group with significantly different survival times. It was observed that patients who fell in the low risk category showed better survival compared to the high risk category (Median survival: undefined versus 42 months; HR: 11.30; P <0.0001). Cox multivariate analysis revealed that risk score was the most significant predictor of survival (HR = 2.587; 95% CI= 1.747-3.832; P= $2.103x10^{-6}$). Thus, the present study suggests that these genes can be used as prognosticators. Risk score which combines the prognostic capability of AP2-α, CHI3L1 and PBEF1 is a much better prognosticator and may help in the selection of high risk cancer patients for tailored treatments.

Key words- AP2-α, CHI3L1, PBEF1, risk score breast cancer, prognosis.

I. INTRODUCTION

Breast cancer is the most common malignancy prevailing among women population across the world, and if it is diagnosed at an early stage of the disease, its treatment becomes possible (1, 2). Our understanding of tumor biology in recent years has achieved a significant improvement by immunohistochemistry. Evaluation of protein expression in sporadic cancers by immunohistochemistry has enabled the identification of new biomarkers that have diagnostic, therapeutic and prognostic value (3). Prognostic markers are indicators of aggressiveness, invasiveness, extent of spread of tumors, cell proliferation and thus, correlate with survival independent of systemic therapy and can be used to select patients at risk (4). Clinical prognostic factors are axillary lymph node status, tumor size, nuclear grade, histological grade, stage of the disease and distant metastasis. Candidate prognostic biomarkers in breast cancer include elevated levels of estrogen (ER) and progesterone receptor (PR); amplification and over-expression of ErbB2 (5, 6). Review of literature suggests that over expression of ErbB2, AP-2α, AP-2γ and CHI3L1 have prognostic significance in breast cancer, whereas not much is known about PBEF1, and these genes are the focus of our study. AP-2 transcription factors are a family of five developmentally regulated proteins, AP-2α, β, γ , δ and θ, encoded by five different genes (7, 8). AP-2 proteins play an important role as regulators of gene expression in development, cell growth, cell differentiation and apoptosis. Alterations in AP-2 function have been linked with malignancy and studies suggest that AP-2 family acts as tumor suppressor genes (9). In breast and other cancers, reduced expression of AP-2 *in vivo* has been linked with disease progression (10-14). The AP-2 gene family has been implicated in regulation of both ER and ERB2 expression during the progression from normal breast epithelium to breast cancer (15), and represents a clinical marker of poor prognosis in breast cancer. CHI3L1 is a 40-kDa secreted glycoprotein that was discovered as a heparin-binding protein in the medium of human synoviocytes, chondrocytes, and MG-63 osteosarcoma cell line (16-18). CHI3L1may play a role in the proliferation, differentiation, invasiveness, apoptosis, angiogenesis, remodeling of the extracellular matrix and stimulation of fibroblasts surrounding the tumor (19). Thus, CHI3L1expression is associated with poor prognosis and shorter disease free survival in breast cancer and other malignancies (20-25). Therefore, CHI3L1can be termed as a biomarker that can be used as a prognosticator. PBEF1 functions as an enzyme involved in NAD biosynthesis (26). PBEF1 has also been recognized as a hormone named Visfatin as it was found in high levels in visceral fat (27). PBEF1 has been shown to be over expressed in colon, pancreatic, colorectal, astrocytic cancers. However, the role of PBEF1 in breast cancer is largely not known (28, 29). In view of this and with emerging data on the role

of some of these prognostic markers in breast cancer, the aim of this study was to evaluate the association of four genes (AP-2α, AP-2γ, CHI3L1 and PBEF1) with survival of patients with breast cancer. To accomplish this, we generated risk score, which combined the independent capability of AP-2α, CHI3L1 and PBEF1 to predict prognosis.

II. METHODS

Patient population.

The retrospective study included 174 breast cancer patients who underwent surgical resection at Kidwai Memorial Institute of Oncology between 2001 and 2002. The study protocol was approved by the Ethics Committee of Kidwai Memorial Institute of Oncology, Bangalore India. The patients were followed up at regular intervals and their clinical status was documented. Patients who met the following eligibility criteria were included: (1) diagnosis of breast tumors identified by histopathological examination; (2) all patients who underwent surgery (modified radical mastectomy, lumpectomy) followed by adjuvant radiation therapy and chemotherapy. All patients who were estrogen receptor positive had also undergone tamoxifen therapy. (3) Availability of follow-up data; (4) no preoperative treatment, such as chemotherapy and radiotherapy; (5) no history of familial malignancy or other synchronous malignancy (such as ovarian tumor); and (6) no death in the perioperative period. Overall survival was defined as the duration between surgery and death of the patient due to disease or the last follow up. Patient characteristics are listed in Table 1.

Tissue specimens.

Paraffin-embedded breast cancer samples were obtained from Kidwai Memorial Institute of Oncology. The samples were collected between 2001 and 2002 from patients who underwent surgical resection. Each tumor sample was assigned a histological grade based on the World Health Organization (WHO) classification criteria. clinical stage of the disease based on AJCC, 2002; distant metastasis based on TNM classification of breast tumors by UICC, 2002; lymph node status classified based on UICC, 2002; tumor size, classified as T1-4 based on TNM classification of breast tumors by UICC, 2002; pathological grade of the tumor based on Elston and Ellis, 1991; histological type of tumor based on WHO classification of tumors, 2003.

Immunohistochemistry.

IHC for AP-2α, AP-2γ, CHI3L1 and PBEF1 was done using AP-2α polyclonal antibody (C-18; Cat# sc-184; Santa Cruz Biotechnology, USA), AP-2γ polyclonal antibody (6E4/4; Cat# sc-12762; Santa Cruz Biotechnology, USA), PBEF1 antibody (Rabbit polyclonal antibody against purified GST-PBEF1 protein was made using standard immunization protocol) and CHI3L1 antibody (Rabbit polyclonal antibody against purified GST-CHI3L1 protein was made using standard immunization protocol) respectively. All samples were fixed in 10% formalin and paraffin embedded. Serial 3-5μm sections were cut from the retrieved cases and stained with Haematoxylin and Eosin to confirm the presence of tumor. Tissue sections were incubated overnight at 60ºC, deparaffinized for 15 minutes two times in xylene. Hydration in absolute alcohol was carried out followed by rinsing in distilled water. Using Citrate buffer, pH 6.0, antigen retrieval was performed in a pressure cooker at high pressure for 20-30 minutes. The endogenous peroxidases were blocked by 0.5% hydrogen peroxide treatment for 20 minutes and the antibody blocking was carried out by immersing the slides in 2% milk for 30 minutes. Sections were incubated for 90 minutes at 24ºC with mouse anti-human monoclonal AP-2α antibody at a dilution of 1:100, with mouse anti-human monoclonal AP-2γ antibody at a dilution of 1:50, PBEF antibody at a dilution of 1:500 and CHI3L1antibody at a dilution of 1:500. After rinsing with TBS buffer, the sections were incubated with Biotin-2º antibody (BioGenex, SanRamon, CA) for 30 minutes at 24ºC. Finally Streptavidin-3º antibody (BioGenex, SanRamon, CA) complex was added and sections were incubated at 24ºC for 30 minutes. Visualization of the antigen-antibody complex was achieved by using the Diaminobenzidine tetrahydrochloride detection system (BioGenex, SanRamon, CA). After final washing, slides were counterstained with Mayer's hematoxylin, dehydrated and mounted. For AP2α, a glioblastoma sample served as a positive control [11]. In addition, there are other studies which have used the same antibody [9, 10, 31, 32, 33]. For PBEF1 also, a glioblastoma sample served as a positive control [29]. For AP-2γ immunostaining, the antibody used in this study has been used previously [9,31.32.33.35]. The specificity of CHI3L1 antibody was confirmed by western blot analysis that detected the protein of expected size from cell line extract which served as a positive control- data not shown. Negative controls included exclusion of the 1º antibody and replacement with TBS buffer. Estrogen receptor (ER), Progesterone receptor (PR) staining and ErbB-2 staining details are furnished in our manuscript (30).

Scoring of Immunoreactivity.

Tumor epithelial cellular immunostaining for AP-2α, AP-2γ, and PBEF1and CHI3L1 were assessed in test sections by a pathologist. Labeling index was calculated based on number of cells positive and the intensity of staining. Based on cell positivity the labeling index ranged from <10%-100% and the scores were scaled as 0 if no staining was present. Score 1 was used if ≤ 10% of cells

were stained, score 2 if 10-50%, score 3 if 51-80% and score 4 if more than 80% of cells showed immunohistochemical staining reaction. Based on cell intensity, the values given are mild=1, moderate=2 and severe=3. Labeling index which is a measure of expression ranged from 0 to 12 [35]. Control slides were checked for non-specific binding before assessing the staining intensity and percentage positivity of the tumor cells in test sections.

Survival and Statistical analysis.

Patients for whom expression details of AP-2α, AP-2γ, CHI3L1 and PBEF1 were available were included for analysis. Mean age of the patient cohort was 45 years. The follow up period was between January 2001 to January 2012, with the minimum and maximum follow up being 12 months and 132 months respectively. The mean survival period was (Median: 54; Range: 12 months to 10 years).

The correlation of expression of a given gene with survival was assessed by Cox regression method. The formula was devised using the Cox regression coefficients derived from the Cox proportional hazard analysis. Each patient was assigned a risk score that is a linear combination of the expression levels of the significant parameters weighted by their respective Cox regression coefficients. According to our analysis, patients having high risk scores are expected to have poor survival outcomes as compared to patients having low risk scores. The study involves calculation of risk scores for every patient as follows:

Risk score = ((-0.705 X labeling index of AP2-α) + (0.300 X labeling index of CHI3L1) + (0.279 X labeling index of PBEF1))

The significant parameters that formed the signature were of two types - risky and protective. Risky parameters were defined as those that had hazard ratio greater than 0. Protective parameters were defined as those that had hazard ratio lesser than 0. Using this definition, we found 1 protective parameter and 6 risky parameters.

Based on the risk score, the patients were divided into high-risk and low-risk groups using the median risk score as the cut-off. We used Cox regression analysis to evaluate the contribution of AP-2α, CHI3L1, PBEF1, ErbB2, stage of the disease, distant metastasis, lymph node status, estrogen receptor, progesterone receptor, tumour size, grade, histological type, menopausal status, age, AP-γ and risk score as prognostic factors. Multivariate analysis was carried out for variables which showed correlation with survival using univariate analysis. Kaplan-Meier method was used to estimate overall survival. Graph Pad Prism 5.0 software was used for Kaplan-Meier graph plotting and calculation of P-values. P-values less than 0.05 were considered significant.

III. RESULTS

Expression pattern of AP2α, AP2-γ, PBEF1 and CHI3L1.

We detected levels of AP-2α, AP2-γ, CHI3L1 and PBEF1 by immunohistochemistry (IHC). Breast cancer tissues showed predominant nuclear and lower cytoplasmic staining for AP-2α (**Figure 1B**); nuclear staining for AP2-γ (**Figure 1D**); cytoplasmic staining for CHI3L1 (**Figure 1F**) and PBEF1 (**Figure 1H**). Whereas, adjacent normal breast cells showed epithelial nuclear and myoepithelial nuclear staining for AP-2α (**Figure 1A**) and AP2-γ (**Figure 1C**) respectively; epithelial cytoplasmic staining for CHI3L1 (**Figure 1E**) and PBEF1 (**Figure 1G**). For statistical analysis, the samples were divided into low or high expression groups using median labeling index (LI) as cut off. LI was calculated based on number of cells positive and the intensity of staining and it ranged from 0-12 for all the three genes. Out of 104 samples available, 46% (48/104) cases were found to show high expression (LI of ≥4 (median of LI) was considered as high expression) for AP-2α (nuclear) immuno staining; 56% (96/172) for CHI3L1 (LI of ≥4 (median of LI) was considered as high expression) and 46% (50/109) for PBEF1 (LI of ≥9 (median of LI) was considered as high expression).

*Correlation of different markers with overall survival by univariate Cox regression analysis***.**

The different markers were subjected to univariate Cox proportional hazard regression analysis. AP2-α, CHI3L1, PBEF1, ErbB2, Stage of the disease, distant metastasis, lymph node involvement, and tumor size correlated with patient survival (Table 2). AP-2α (nuclear) was seen to be a good prognostic marker, while CHI3L1 and PBEF1 were found to be poor prognostic markers. However, parameters like AP-2γ, estrogen receptor, progesterone receptor, grade, histological subtype, menopausal status and age of the patient showed no significant correlation with patient survival.

AP-2α, CHI3L1 and PBEF1 are independent predictors of overall survival

The markers which stood significant in univariate analysis were subjected to forward condition multivariate Cox proportional hazard regression analysis. We found that AP2-α (HR = 0.609; B = -0.495; p = 0.005), CHI3L1 (HR = 1.361; B = 0.308; p = 0.001), PBEF1 (HR = 1.541; B = 0.433; p = 0.004), and confonding factor Stage (HR = 3.207; B = 1.165; p = 0.03) were independent

predictors of survival in breast cancer patients (Table 3). However, ErbB2, distant metastasis and lymph node lost their significance in multivariate analysis.

Stratification of patients based on expression of AP2-α, CHI3L1 and PBEF1:

With the results of univariate and multivariate analysis indicating the reliability of AP2-α, CHI3L1, and PBEF1 to predict survival, we attempted stratifying the patients into high or low risk groups based on the expression. Patients with tumors having high expression level of AP-2α nuclear positivity had better survival than patients expressing low levels (median survival: undefined versus 76 months; HR: 7.210; P <0.0001) (**Figure 2A**) suggesting that it is a good prognostic marker, whereas patients having low expression levels of CHI3L1 and PBEF1 had better survival than patients expressing high levels (median survival: undefined versus 75 months; HR: 5.442; P <0.0001 and median survival: undefined versus 53 months; HR: 4.929; P <0.0001 respectively) (**Figure 2B and 2C respectively**) which indicates that these are poor prognostic markers.

Risk score based on AP2α, CHI3L1and PBEF1is an independent predictor of survival**.**

To combine the prognostic capabilities of the three genes (AP2-α, CHI3L1 and PBEF1) of our interest, we calculated a risk score that could predict patient survival (**Table 4)**. Risk score is a linear combination of the expression levels of the significant parameters weighted by their respective Cox regression coefficients. According to our analysis, patients having high risk scores are expected to have poor survival outcomes as compared to patients having low risk scores. The risk score in the cohort ranged from -7.56 to 6.948, with the median value of 1.857. The risk score was evaluated for its influence on survival of the patients and was found to significantly correlate with survival as a continuous variable (HR= 2.532; P = 2.912×10^{-8}). To assess the contribution and independency of risk score in predicting survival, multivariate analysis was carried out. ErbB2, stage, distant metastasis and nodal involvement were used in multivariate to adjust the risk score for independency. Risk score was seen to highly correlate with survival (HR = 2.587; p = 2.103×10^{-6}) and was independent of ErbB2, stage, distant metastasis and nodal involvement (**Table 4**). Stage (HR = 2.260; p = 0.02) also appeared to be independent predictor of survival.

Stratification of patients based on risk score.

It is evident from the above results that risk score is much better than other parameters in predicting survival. Therefore, we attempted to stratify the patients based on risk score to predict survival. The relative expression values of individual genes vary based on the methodologies used, which in turn reflect on the risk score. The range of risk score of the present study cohort ranged from -7.56 to 6.948. The median value of the risk score for the present study cohort was calculated, which was noted to be 1.857. The median cut off divided the patient set into two groups, low risk and high risk. The low risk group showed better survival compared to the high risk group (median survival: undefined versus 42 months; HR: 11.30; P <0.0001) (**Figure 3A**).
Further, to identify patients with very high risk for better treatment, we divided the patient set into 3 groups. While the low risk group remained the same (lesser than 50% i.e. median cut off), at 75% cut off the high risk group (from the previous stratification) was further divided into intermediate and high risk groups. Low and intermediate risk group showed better survival compared to high risk group (median survival: undefined versus 19 months respectively; P <0.0001) (**Figure 3B**). Also it was noted that most of the dead patients belonged to the high and intermediate risk groups, while all those belonging to low risk group were alive at the end of follow up (**Figure 3C**). Patients belonging to low risk group had low risk score and high risk group had high risk score (**Figure 3D**).

Nature of genes involved in risk score analysis.

Of the three genes that we took a closer look at, the nature of one gene was seen to be different from the other two genes. When patients were ranked according to risk score, AP-2α appears to behave protective while CHI3L1 and PBEF were found to be risky with respect to their association between their expression and patient survival. The protective genes were expressed at a higher level in the low risk group compared to the high risk group and the risky genes were expressed at a higher level in the high risk group than in the low risk group (**Table 5, Figure 3E**).

IV. Discussion

In this study, we analyzed the expression pattern of AP–2α, AP-2γ, CHI3L1, PBEF1 and ErbB2 expression among Indian sporadic breast cancer patients by immunohistochemistry and correlated the expression level of these markers with overall survival. Labeling index of the three genes of our interest were used to calculate the risk score, which stood significant in univariate analysis. These genes have been shown to correlate with survival (except AP-2γ). Hence, we combined the prognostic capabilityof AP–2α, CHI3L1 and PBEF1 and generated a risk score which was independent when compared to other known predicting clinical parameters like ErBb2, nodal involvement, stage of the disease and distant metastasis. Patients with high risk score had shorter survival compare to patients with low risk score value. In multivariate analysis risk score was seen to highly correlate with patient survival and was independent of ErbB2, stage, distant metastasis and nodal involvement. However ErbB2, nodal status and metastasis were also shown

to be independent predictors of survival. The three prognostic markers used in this study included one gene (AP-2α) that was protective and two genes (CHI3L1, PBEF1) that were risky with respect to their level in the low risk group compared to the high risk group. Our findings suggest that the protective and risky nature of these genes is suggestive of their function being either inhibitory or promoting respectively of various properties of cancer cells like proliferation, migration and invasion.

In our study, though we have observed nuclear and cytoplasmic staining of AP-2α tumor cells, it was predominantly nuclear staining. While our study is consistent with some (31-34), there are other studies that report only nuclear staining (35, 36). The adjacent normal breast cells show epithelial nuclear cell staining which is consistent with the reported studies (35, 37). Of 104 patients, 46% showed high expression of AP-2α . Patients expressing higher levels (n=48) of AP-2α had survived better than patients expressing lower levels (n=56) (undefined vs 76 months; p = <0.0001). AP-2α has shown to be significant in both univariate and multivariate analysis in our findings, which is consistent with the reported literature (31). AP-2α expression has been shown to be associated with good prognosis in many breast cancer studies (31, 32, 34, 37). Our study shows that AP-2α (nuclear) is a marker of good prognosis in breast cancer patients.

Cytoplasmic staining was observed for CHI3L1. However, other studies have reported only cytoplasmic and not membranous staining (38-40). We observed cytoplasmic epithelial cells staining in adjacent an normal cell which is consistent with Roslind et al., 2007 (39) who has reported epithelial cytoplasmic staining. However, they have reported dot like staining pattern in the nuclear cells which was not seen in our findings. Of 172 patients, 47% showed high expression of CHI3L1. Patients expressing lower levels (n=76) of CHI3L1 had survived better than patients expressing higher levels (n=96) (undefined vs 75 months; p = <0.0001). CHI3L1has shown to be significant in both univariate and multivariate analysis in our findings, which is consistent with the reported literature (38). CHI3L1 expression has been shown to be associated with poor prognosis in many malignancies including breast cancer (20-22, 24, 25, 38, 39), while some studies have also reported that CHI3L1 is not a predictor of survival in breast cancer patients (40). Our study shows that CHI3L1 is a marker of poor prognosis in breast cancer patients.

We have correlated the expression of PBEF1 with survival on a retrospective data of 109 breast cancer patients. Patients expressing lower levels (n=59) of PBEF1 had survived better than patients expressing higher levels (n=50) (undefined vs 53 months; p = <0.0001).We observed cytoplasmic staining pattern for PBEF1 in tumor cells, while adjacent normal cells show cytoplasmic staining in the epithelial layer of the tissue. PBEF1 has shown to be significantly associated with survival in both univariate and multivariate analysis in our findings. Prognostic significance of PBEF1 has not been reported previously for breast cancer and the present study is first of its kind to show that high expression of PBEF1 is significantly associated with shorter survival among breast cancer patients.

We found that these three genes (AP−2α, CHI3L1 and PBEF1) are independent prognostic indicators. Risk score was calculated by combining the individual prognostic capabilities of these three genes. Risk score highly correlated with survival. It was also shown to be an independent prognostic indicator and was a better indicator than other known prognostic markers. When patients were ranked according to risk score, patients with high risk score fell in the high risk category and those with low risk score fell in the low risk category. Hence, risk score can be used to categorize patients. AP-2α appears to behave like a protective gene This is well documented and reported in literature. Gee et al, Friedrichs and Pellikainen et al. suggest that AP-2α is involved in development, apoptosis and cell cycle regulations. Altered expression of this protein is been associated with poor clinical outcome of breast cancer patients. CHI3L1 and PBEF1 seem to be risky genes with respect to their level in the low risk group compared to the high risk group. It has been shown that lectin binding is linked to tumor progression and K-ras activation in colorectal cancer. Hence, CHI3L1 maybe playing a role in tumor invasion [20], which supports our data that it is a marker of poor prognosis. Overexpression of PBEF1 has been correlated with poor response to doxorubicin-based primary chemotherapy in breast cancer [29]. Also, it has been shown to be a poor prognostic indicator in GBM [29], which is in corroboration with our study. Therefore, this study supported by previous studies suggests that AP-2α, CHI3L1 and PBEF1 could be involved in inhibiting or promoting various properties of cancer cells like proliferation, migration and invasion that requires in-depth study. The effect of various drugs on the expression and effect of downregulation/overexpression of these genes on breast tumor growth needs to be explored.

Acknowledgement

This study was supported by a grant from DBT, Government of India to KS and GM. Infrastructural support by funding from ICMR, DBT, DST and UGC to MCB is acknowledged.

References

[1] Andreopoulou E, Hortobagyi GN. Prognostic factors in metastatic breast cancer: successes and challenges toward individualized therapy. J Clin Oncol. 2008 Aug 1;26(22):3660-2.
[2] Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin. 2009 Jul-Aug;59(4):225-49.
[3] Osin PP, Lakhani SR. The pathology of familial breast cancer: Immunohistochemistry and molecular analysis. Breast Cancer Res. 1999;1(1):36-40.

[4] Masood S. Prognostic/predictive factors in breast cancer. Clin Lab Med. 2005 Dec;25(4):809-25, viii.
[5] Taneja P, Maglic D, Kai F, Zhu S, Kendig RD, Fry EA, et al. Classical and Novel Prognostic Markers for Breast Cancer and their Clinical Significance. Clin Med Insights Oncol.4:15-34.
[6] Yamashita H, Nishio M, Toyama T, Sugiura H, Zhang Z, Kobayashi S, et al. Coexistence of HER2 over-expression and p53 protein accumulation is a strong prognostic molecular marker in breast cancer. Breast Cancer Res. 2004;6(1):R24-30.
[7] Hilger-Eversheim K, Moser M, Schorle H, Buettner R. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. Gene. 2000 Dec 30;260(1-2):1-12.
[8] Orso F, Penna E, Cimino D, Astanina E, Maione F, Valdembri D, et al. AP-2alpha and AP-2gamma regulate tumor progression via specific genetic programs. FASEB J. 2008 Aug;22(8):2702-14.
[9] Turner BC, Zhang J, Gumbs AA, Maher MG, Kaplan L, Carter D, et al. Expression of AP-2 transcription factors in human breast cancer correlates with the regulation of multiple growth factor signalling pathways. Cancer Res. 1998 Dec 1;58(23):5466-72.
[10] Anttila MA, Kellokoski JK, Moisio KI, Mitchell PJ, Saarikoski S, Syrjanen K, et al. Expression of transcription factor AP-2alpha predicts survival in epithelial ovarian cancer. Br J Cancer. 2000 Jun;82(12):1974-83.
[11] Britto R, Umesh S, Hegde AS, Hegde S, Santosh V, Chandramouli BA, et al. Shift in AP-2alpha localization characterizes astrocytoma progression. Cancer Biol Ther. 2007 Mar;6(3):413-8.
[12] Hietala KA, Kosma VM, Syrjanen KJ, Syrjanen SM, Kellokoski JK. Correlation of MIB-1 antigen expression with transcription factors Skn-1, Oct-1, AP-2, and HPV type in cervical intraepithelial neoplasia. J Pathol. 1997 Nov;183(3):305-10.
[13] Karjalainen JM, Kellokoski JK, Eskelinen MJ, Alhava EM, Kosma VM. Downregulation of transcription factor AP-2 predicts poor survival in stage I cutaneous malignant melanoma. J Clin Oncol. 1998 Nov;16(11):3584-91.
[14] Lipponen P, Aaltomaa S, Kellokoski J, Ala-Opas M, Kosma V. Expression of activator protein 2 in prostate cancer is related to tumor differentiation and cell proliferation. Eur Urol. 2000 May;37(5):573-8.
[15] Bosher JM, Williams T, Hurst HC. The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma. Proc Natl Acad Sci U S A. 1995 Jan 31;92(3):744-7.
[16] Hakala BE, White C, Recklies AD. Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family. J Biol Chem. 1993 Dec 5;268(34):25803-10.
[17] Johansen JS, Williamson MK, Rice JS, Price PA. Identification of proteins secreted by human osteoblastic cells in culture. J Bone Miner Res. 1992 May;7(5):501-12.
[18] Shackelton LM, Mann DM, Millis AJ. Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodeling. J Biol Chem. 1995 Jun 2;270(22):13076-83.
[19] Shao R, Hamel K, Petersen L, Cao QJ, Arenas RB, Bigelow C, et al. YKL-40, a secreted glycoprotein, promotes tumor angiogenesis. Oncogene. 2009 Dec 17;28(50):4456-68.
[20] Cintin C, Johansen JS, Christensen IJ, Price PA, Sorensen S, Nielsen HJ. Serum YKL-40 and colorectal cancer. Br J Cancer. 1999 Mar;79(9-10):1494-9.
[21] Dupont J, Tanwar MK, Thaler HT, Fleisher M, Kauff N, Hensley ML, et al. Early detection and prognosis of ovarian cancer using serum YKL-40. J Clin Oncol. 2004 Aug 15;22(16):3330-9.
[22] Johansen JS, Drivsholm L, Price PA, Christensen IJ. High serum YKL-40 level in patients with small cell lung cancer is related to early death. Lung Cancer. 2004 Dec;46(3):333-40.
[23] Johansen JS, Jensen BV, Roslind A, Nielsen D, Price PA. Serum YKL-40, a new prognostic biomarker in cancer patients? Cancer Epidemiol Biomarkers Prev. 2006 Feb;15(2):194-202.
[24] Schmidt H, Johansen JS, Gehl J, Geertsen PF, Fode K, von der Maase H. Elevated serum level of YKL-40 is an independent prognostic factor for poor survival in patients with metastatic melanoma. Cancer. 2006 Mar 1;106(5):1130-9.
[25] Tanwar MK, Gilbert MR, Holland EC. Gene expression microarray analysis reveals YKL-40 to be a potential serum marker for malignant character in human glioma. Cancer Res. 2002 Aug 1;62(15):4364-8.
[26] Samal B, Sun Y, Stearns G, Xie C, Suggs S, McNiece I. Cloning and characterization of the cDNA encoding a novel human pre-B-cell colony-enhancing factor. Mol Cell Biol. 1994 Feb;14(2):1431-7.
[27] Pilz S, Mangge H, Obermayer-Pietsch B, Marz W. Visfatin/pre-B-cell colony-enhancing factor: a protein with various suggested functions. J Endocrinol Invest. 2007 Feb;30(2):138-44.
[28] Hufton SE, Moerkerk PT, Brandwijk R, de Bruine AP, Arends JW, Hoogenboom HR. A profile of differentially expressed genes in primary colorectal cancer using suppression subtractive hybridization. FEBS Lett. 1999 Dec 10;463(1-2):77-82.
[29] Reddy PS, Umesh S, Thota B, Tandon A, Pandey P, Hegde AS, et al. PBEF1/NAmPRTase/Visfatin: a potential malignant astrocytoma/glioblastoma serum marker with prognostic value. Cancer Biol Ther. 2008 May;7(5):663-8.
[30] Vaidyanathan K, Kumar P, Reddy CO, Deshmane V, Somasundaram K, Mukherjee G. ErbB-2 expression and its association with other biological parameters of breast cancer among Indian women. Indian J Cancer. Jan-Mar;47(1):8-15.
[31] Pellikainen J, Kataja V, Ropponen K, Kellokoski J, Pietilainen T, Bohm J, et al. Reduced nuclear expression of transcription factor AP-2 associates with aggressive breast cancer. Clin Cancer Res. 2002 Nov;8(11):3487-95.
[32] Pellikainen J, Naukkarinen A, Ropponen K, Rummukainen J, Kataja V, Kellokoski J, et al. Expression of HER2 and its association with AP-2 in breast cancer. Eur J Cancer. 2004 Jul;40(10):1485-95.

[33] Pellikainen JM, Ropponen KM, Kataja VV, Kellokoski JK, Eskelinen MJ, Kosma VM. Expression of matrix metalloproteinase (MMP)-2 and MMP-9 in breast cancer with a special reference to activator protein-2, HER2, and prognosis. Clin Cancer Res. 2004 Nov 15;10(22):7621-8.
[34] Allouche A, Nolens G, Tancredi A, Delacroix L, Mardaga J, Fridman V, et al. The combined immunodetection of AP-2alpha and YY1 transcription factors is associated with ERBB2 gene overexpression in primary breast tumors. Breast Cancer Res. 2008;10(1):R9.
[35] Friedrichs N, Jager R, Paggen E, Rudlowski C, Merkelbach-Bruse S, Schorle H, et al. Distinct spatial expression patterns of AP-2alpha and AP-2gamma in non-neoplastic human breast and breast cancer. Mod Pathol. 2005 Mar;18(3):431-8.
[36] Gee JM, Eloranta JJ, Ibbitt JC, Robertson JF, Ellis IO, Williams T, et al. Overexpression of TFAP2C in invasive breast cancer correlates with a poorer response to anti-hormone therapy and reduced patient survival. J Pathol. 2009 Jan;217(1):32-41.
[37] Gee JM, Robertson JF, Ellis IO, Nicholson RI, Hurst HC. Immunohistochemical analysis reveals a tumour suppressor-like role for the transcription factor AP-2 in invasive breast cancer. J Pathol. 1999 Dec;189(4):514-20.
[38] Kim SH, Das K, Noreen S, Coffman F, Hameed M. Prognostic implications of immunohistochemically detected YKL-40 expression in breast cancer. World J Surg Oncol. 2007;5:17.
[39] Roslind A, Johansen JS, Junker N, Nielsen DL, Dzaferi H, Price PA, et al. YKL-40 expression in benign and malignant lesions of the breast: a methodologic study. Appl Immunohistochem Mol Morphol. 2007 Dec;15(4):371-81.
[40] Roslind A, Knoop AS, Jensen MB, Johansen JS, Nielsen DL, Price PA, et al. YKL-40 protein expression is not a prognostic marker in patients with primary breast cancer. Breast Cancer Res Treat. 2008 Nov;112(2):275-85.

AUTHORS

First Author – Umaira Tabassum, Ph.D research scholar, Department of Pathology, Kidwai Memorial Institute of Oncology, Bangalore 68. E-mail: umairatabu@gmail.com
Second Author –Irene Rosita Pia Patric, Research associate, Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 12. E-mail: piapatrick@yahoo.com
Third Author – Dr Geetashree Mukherjee, Prof & Head , Department of Pathology, Kidwai Memorial Institute of Oncology, Bangalore 68. E-mail: mukherjeegeeta@yahoo.co.in
Fourth Author- Dr Kumaravel Somasundaram, Associate professor**,** Microbiology and Cell Biology, Indian Institute of Science, Bangalore 12. E-mail: skumar@mcbl.iisc.ernet.in

Correspondence Author – Dr Geetashree Mukherjee, Prof & Head , Department of Pathology, Kidwai Memorial Institute of Oncology, Bangalore 68. E-mail: mukherjeegeeta@yahoo.co.in/ kidwaipathology1@gmail.com
Tel: 080 26094000, Fax: 080 26560723

Table 1. Patient characteristics

Characteristic	**Number of patients (%)**
Total number of patients	174
AP-2α nuclear (n=104)	
High expression	48 (46 %)
Low expression	56 (54 %)
AP-2α cytoplasmic (n=104)	
High expression	49 (47%)
Low expression	55 (53%)
AP-2γ (n=72)	
High expression	33 (46%)
Low expression	39 (54%)
CHIL31 (172)	
High expression	96 (56%)
Low expression	76 (42 %)
PBEF1 (109)	
High expression	50 (46 %)

Low expression	59 (54 %)
ErbB2 (174)	
High expression	83 (48%)
Low expression	91 (52%)
Age (n=174)	
</=45	93 (53%)
>45	81 (47%)
Tumor size (n=160)	
T1	13 (8%)
T2	78 (49%)
T3	32 (20%)
T4	37 (23%)
Lymph node status (n=170)	
Positive	109 (64%)
Negative	61 (36%)
Distant metastasis (n=169)	
Positive	14 (9%)
Negative	155 (91%)
Stage (n=159)	
1	6 (4%)
2	52 (33%)
3	87 (55%)
4	14 (8%)
Histological type (n=164)	
Ductal	157 (96%)
Lobular	7 (4%)
Histological grade (n=158)	
1	2 (2%)
2	19 (12%)
3	137 (86%)
ER status (n=174)	
Positive	100 (57%)
Negative	74 (43%)
PR status (n=174)	
Positive	96 (55%)
Negative	78 (45%)
Menopausal status (n=168)	
Premenopausal	82 (48%)
Postmenopausal	86 (52%)

'n' indicates the number of patients for whom information was available

Table 2. Univariate Cox regression analysis

Variable	Regression coefficient	Hazard ratio (95% CI)	P-Value

AP-2α nuclear	-0.705	0. 494 (0.346-0.706)	0.0001
AP-2α cytoplasmic	0.146	1. 157 (1.027-1.303)	0.017
AP-2γ	-0.180	0.835 (0.417-1.673)	0.611
CHIL31	0.300	1. 350 (1.238-1.473)	$2.443x10^{-11}$
PBEF1	0.279	1. 322 (1.143-1.529)	0.0001
ErbB2	1.302	3.677 (1.991-6.794)	$3.206x10^{-5}$
Stage[a]	1.474	4.368 (2.659-7.714)	$5.770x10^{-9}$
Distant metastasis[b]	1.850	6.360 (3.346-12.091)	$1.656x10^{-8}$
Lymph node status[c]	1.194	3.330 (1.563-7.095)	0.002
Estrogen receptor	-0.106	0.899 (0.520-1.555)	0.704
Progesterone receptor	-0.093	0.911 (0.528-1.572)	0.738
Tumor size[d]	0.493	1.637 (1.193-2.246)	0.002
Grade[e]	0.155	1.167 (0.545-2.501)	0.691
Histological type[f]	1.020	2.772 (0.383-20.086)	0.313
Menopausal status	-0.062	0.940 (0.540-1.637)	0.826
Age	-0.009	0.991 (0.967-1.016)	0.475

a, clinical stage of the disease based on AJCC, 2002; **b**, distant metastasis based on TNM classification of breast tumors by UICC, 2002; **c**, lymph node status classified based on UICC, 2002; **d**, tumor size, classified as T 1-4 based on TNM classification of breast tumors by UICC, 2002; **e**, pathological grade of the tumor based on Elston and Ellis, 1991; **f**, histological type of tumor based on WHO classification of tumors, 2003

Table 3. Multivariate Cox regression analysis

Variable	**Regression coefficient**	**Hazard ratio (95% CI)**	**P-Value**
AP-2α nuclear	-0.495	0. 609 (0.430-0.863)	0.005
CHI3L1	0.308	1. 361 (1.139-1.627)	0.001
PBEF1	0.433	1.541 (1.144-2.076)	0.004
Stage	1.165	3.207 (1.122-9.165)	0.03

Table 4. Prognostic value of risk score – Univariate and multivariate analysis

Variable	**Hazard ratio (95% CI)**	**P-Value**
I-Univariate analysis		
Risk score	2.532 (1.823-3.516)	$2.912x10^{-8}$

II-Multivariate analysis		
Risk score	2.587 (1.747-3.832)	$2.103x10^{-6}$
Stage	2.260 (1.125-4.541)	0.02

Table 5. Expression of AP-2*α*, CHI3L1 and PBEF1 in high and low risk group

Gene	**Low risk**		**High risk**	
	Median	**Mean ± SD**	**Median**	**Mean ± SD**
AP-2α	8.000	8.764 ± 2.177	1.000	1.188 ± 1.330
CHI3L1	2.000	2.052 ± 1.305	9.000	9.520 ±2.466
PBEF1	4.000	3.934 ± 2.901	12.000	11.380 ± 1.223

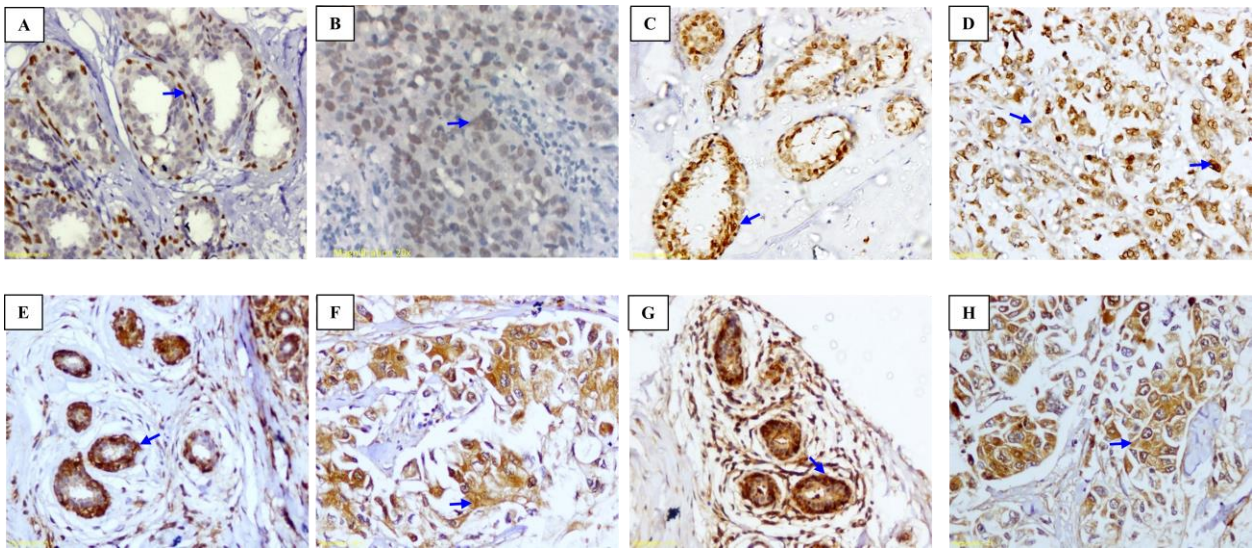

Figure 1: Immunohistochemical staining patterns for AP-2α, AP-2γ, CHI3L1 and PBEF1
The arrows indicate the region of staining. An example of (**A**) AP-2α normal tissue showing epithelial nuclear positive staining, (**B**) AP-2α tumor tissue showing strong cytoplasmic (upper arrow) and nuclear staining (lower arrow), (**C**) AP-2γ normal tissue showing myoepithelial nuclear positive staining, (**D**) AP-2γ tumor tissue showing strong nuclear staining, (**E**) CHI3L1 normal tissue showing epithelial cytoplasmic positive staining, (**F**) CHI3L1 tumor tissue showing strong cytoplasmic staining, (**G**) PBEF1 normal tissue showing epithelial cytoplasmic positive staining and (**H**) PBEF1 tumor tissue showing strong cytoplasmic staining.

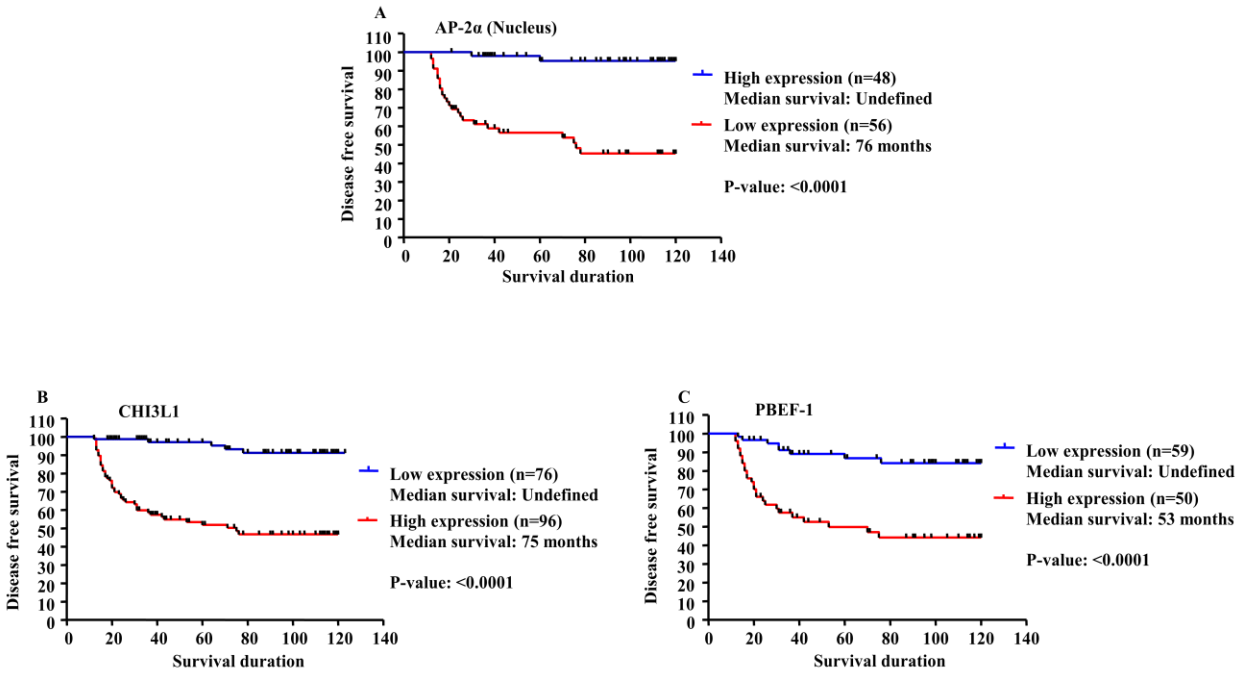

Figure 2: Expression of AP-2α, CHI3L1 and PBEF1 and breast cancer patient survival
Kaplan Meier graphs for A. AP-2α (Nucleus), B. CHI3L1 and C. PBEF1. In all cases, low risk group had better survival than high risk group.

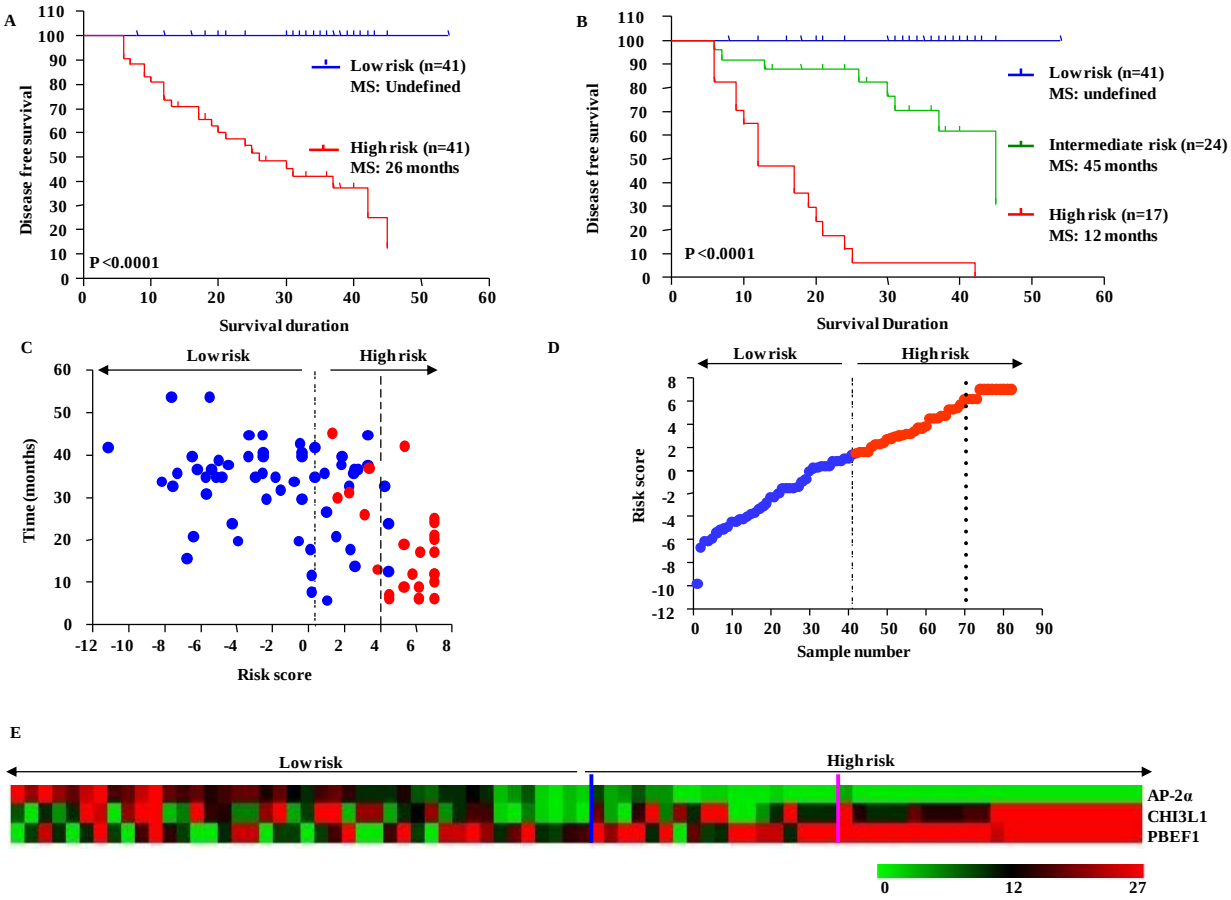

Figure 3: Kaplan-Meier survival estimates of breast cancer patients according to risk score

A. Kaplan Meier graph for risk score analysis showing low and high risk group classification,

B. Kaplan Meier graph for risk score analysis showing low, intermediate and high risk groups,

C. Distribution of patients according to their survival, risk score and censoring status. Red dots indicate patients who have died and blue dots indicate patients who were alive. All dead patients belonged to the high risk group. The dashed line divides the patients into 2 groups- low and high risk. Patients who lie before the dashed line are low risk and after the dotted line are high risk patients of second stratification The patients that lie between the dashed and the dotted line are those who belong to the intermediate risk group of the second stratification,

D. Distribution of patients in the low and high risk groups according to their risk score. The dashed line divides the patients into 2 groups- low and high risk. The patients that lie between the dashed and the dotted line are those who belong to the intermediate risk group.

E. Heat map showing expression pattern of AP-2α (Nucleus), CHI3L1 and PBEF1. The blue line divides the patients into 2 groups- low and high risk. Patients who lie before the blue line are low risk and after the pink line are high risk patients of second stratification The patients that lie between the blue and pink line are those who belong to the intermediate risk group of the second stratification.