"Letter to the editor"

The basic helix-loop-helix (bHLH) proteins in breast cancer progression

Nuzhat N. Kabir\(^1\), Lars Rönnstrand\(^2\) and Julhash U. Kazi\(^{1,2,*}\),

\(^1\)Laboratory of Computational Biochemistry, KN Biomedical Research Institute, Bagura Road, Barisal, Bangladesh,

\(^2\)Experimental Clinical Chemistry, Department of Laboratory Medicine, Lund University, Wallenberg Laboratory, Skåne University Hospital, 20502 Malmö, Sweden.

*Corresponding author:
Julhsah U. Kazi
Experimental Clinical Chemistry
Wallenberg Laboratory, Department of Laboratory Medicine
Lund University, Skåne University Hospital
20502 Malmö, Sweden
E-mail: kazi.uddin@med.lu.se
Tel.: +46 40 33 25 96
Fax: +46 40 33 11 04

Keywords: Basic helix-loop-helix, bHLH, BHLHE40, CDK5RAP3, HES2, HIF1A, ID2, MAX, MLXIP and USF2.
The basic helix-loop-helix (bHLH) proteins are the major transcription factors acting as transcriptional enhancers or inhibitors of various genes through binding to the canonical E-box sequence. The bHLH proteins are highly conserved in both vertebrates and invertebrates. The bHLH domain is approximately 60 amino acids in length consisting of DNA-binding basic region followed by two alpha-helices separated by a variable loop. The DNA-binding basic region associates with the hexanucleotide E-box sequence. The human genome contains 121 bHLH genes which regulate various cellular processes including embryonic development (1,2). Many of them are involved in cell proliferation, differentiation, oncogenesis, as well as apoptosis. For example, MyoD and NeuroD family bHLH proteins contribute to myogenesis, neurogenic differentiation and pancreatic development, while Myc family bHLH proteins are directly involved in cell differentiation, proliferation and oncogenesis (3).

The bHLH proteins are mainly subdivided in six major groups (A-F). This classification is based on the target E-box sequence. While A-C group members target 6 different hexanucleotide sequences, group D members, ID (inhibitor of differentiation/DNA binding) family proteins lack the basic domain and are thus incapable of interacting DNA. The ID family proteins form protein-protein dimers and function as dominant negative regulator of other bHLH proteins. Expression of ID proteins and their target bHLH proteins differs with the cell type and stage of differentiation, growth or development, and balance of bHLH and ID activity is critical to maintain normal cellular process (3). Thus imbalance in these bHLH proteins contributes to abnormal gene regulation, which is very often observed in cancer (3). In this report we studied expression of bHLH proteins in breast cancer.

Micro-array data available for breast cancer patients were downloaded for NCBI gene expression omnibus (GEO). Data were thoroughly checked for platforms and only Affymetrix Human Genome U133A Array (GPL96) was selected for this study. The methods used for data processing were described previously (4-9). Recent studies suggest that several bHLH proteins are involved in breast cancer control by inducing apoptosis or controlling anti-apoptotic genes (10-12). To understand which genes are
deregulated in breast cancer, we analyzed mRNA expression data from more than 2500 patient samples. The complete set of 121 bHLH genes were used for this analysis (1). We used classical multidimensional scaling (MDS) and principal component analysis (PCA) to identify deregulated bHLH genes. The distance matrix was created using "dist" command in R package. Then "cmdscale" command was used to reduce dimensions. This analysis resulted in a list of 14 deregulated genes (Fig. 1A). Then we applied PCA analysis using "prcomp" command on the same dataset. PCA analysis also resulted in a list of 13 deregulated genes (Fig. 1A). We furthermore analyzed dataset with SAM which generated a list of 30 genes (Fig. 1A). Eight genes BHLHE40, CDK5RAP3, HES2, HIF1A, ID2, MAX, MLXIP and USF2 were common in all three experiments which also supported by the heat map of all bHLH genes (Supplementary figure). Thus we conclude that these eight genes are mostly deregulated bHLH genes in breast cancer.

Then we analyzed whether expression of BHLHE40, CDK5RAP3, HES2, HIF1A, ID2, MAX, MLXIP and USF2 genes has a correlation with patient age. We analyzed samples from patients aged 24 years to 93 years. Most of analyzed genes displayed a poor correlation with patient age (Fig. 1B). Furthermore, expression of BHLHE40, CDK5RAP3, HIF1A, ID2, MAX, MLXIP and USF2 was significantly decreased with increased disease grades (Fig. 1C). While expression of BHLHE40, CDK5RAP3, HIF1A, ID2 and MAX was significantly decreased in Black and Hispanic patient groups in compare to Asian patient group, expression of HES2 and USF2 was significantly increased in Hispanic patient group (Fig. 1D). In addition, we also observed a significant increase of BHLHE40, CDK5RAP3, ID2, MAX and USF2 genes expression in estrogen (ER) and progesterone positive patient groups (Fig 1E). Thus we suggest that expression of bHLH genes is deregulated in breast cancer and can be used to classify patients.

Conflict of Interest: The authors declare no conflict of interest.
References:


Figure legend:

Fig. 1: (A) Deregulated bHLH genes in breast cancer. (B) Correlation bHLH genes and breast cancer patient age. (C) Expression of bHLH genes in respect to breast cancer grades. (D) Expression of bHLH genes in Asian, Black and Hispanic races. (E) Expression of bHLH genes in patients with different mutations.
Supplementary figure: Heat map of bHLH genes in breast cancer patients.