Meta Gene 4 (2015) 129–141

Contents lists available at ScienceDirect

Meta Gene

Enriched transcription factor signatures in triple negative breast cancer indicates possible targeted therapies with existing drugs

Scooter Willis [a,*], Pradip De [a], Nandini Dey [a], Bradley Long [b], Brandon Young [a], Joseph A. Sparano [c], Victoria Wang [d], Nancy E. Davidson [e], Brian R. Leyland-Jones [a]

[a] Avera Cancer Institute, Sioux Falls, SD, United States
[b] The Scripps Research Institute, Jupiter, FL, United States
[c] Montefiore Medical Center, Bronx, NY, United States
[d] Dana Farber Cancer Institute, Boston, MA, United States
[e] University of Pittsburgh Cancer Institute and UPMC Cancer Center, Pittsburgh, PA, United States

ARTICLE INFO

Article history:
Received 18 February 2015
Revised 9 April 2015
Accepted 15 April 2015
Available online 15 May 2015

Keywords:
Breast cancer
Triple negative
ER+
GSEA
NR6A1

ABSTRACT

Purpose: Triple negative (TN) breast cancers which lack expression of the estrogen (ER), progesterone (PR), and human epidermal growth factor 2 (HER2) receptors convey a poor prognosis due in part to a lack of targeted therapies.

Methods: To identify viable targets for the treatment of TN disease, we have conducted a gene set enrichment analysis (GSEA) on seven different breast cancer whole genome gene expression cohorts comparing TN vs. ER+ HER2− to identify consistently enriched genes that share a common promoter motif. The seven cohorts were profiled on three different genome expression platforms (Affymetrix, Illumina and RNAseq) consisting in total of 2088 samples with IHC metadata.

Results: GSEA identified enriched gene expression patterns in TN samples that share common promoter motifs associated with SOX9, E2F1, HIF1A, HMGA1, MYC BACH2, CEBPB, and GCNF/NR6A1. Unexpectedly, NR6A1 an orphan nuclear receptor normally expressed in germ cells of gonads is highly expressed in TN and ER+ HER2− samples making it an ideal drug target.

Conclusion: With the increasing number of large sample size breast cancer cohorts, an exploratory analysis of genes that are consistently enriched in TN sharing common promoter motifs allows for the identification of possible therapeutic targets with extensive validation in patient derived data sets.

* Corresponding author.

http://dx.doi.org/10.1016/j.mgene.2015.04.002

Introduction

Breast cancer is the most commonly diagnosed cancer in women (more than 230,000 women were diagnosed with breast cancer in the US last year) and second leading cause of cancer-related deaths, statistics that strongly advocate for a better understanding of the mechanisms that drive mammary carcinogenesis (Siegel et al., 2012). Cancers including breast cancer are initiated as a result of changes that occur in the genome. Differential gene expression analysis is commonly used to reveal the deregulated molecular mechanisms of complex diseases including cancer. Gene expression profiling has classified breast cancer into intrinsic subtypes (Carey et al., 2006; Sørlie et al., 2001). Among these is the basal-like subtype, representing ER/PR-negative with low HER2 expressing tumors are characterized as the triple negative breast cancer (TNBC). Thus, TNBC is defined by histopathologies as a subtype that lacks the expression of estrogen receptor (ER), progesterone receptor (PR), and HER2 amplification/overexpression (Dey et al., 2013a; Rastelli et al., 2010; Foulkes et al., 2010). TNBC accounts for approximately 15% of breast cancer diagnoses, but approximately 25% of breast cancer-related deaths due to a more aggressive biology and lack of targeted therapies (Grigoriadis et al., 2012; Dey et al., 2013b). Gene expression data has documented the heterogeneous nature of TNBC (Mayer et al., 2014).

Despite significant success of targeted anticancer therapies in ER + or HER2 + subtypes of breast cancers, patients with loco-regionally advanced or metastatic triple negative carcinoma have very limited therapy options, especially as chemo-resistance develops to standard chemotherapy. A lack of standardized markers that differentiate “basal-like” and TN subtypes underlines the heterogeneous nature of these cancers. Despite the unclear delineation between TN and “basal-like” they compose some of the worst prognoses in breast cancer, are associated with an undifferentiated metaplastic histology with stem cell-like characteristics and have a high incidence of metastasis (Lien et al., 2007; Ben-Porath et al., 2008; Honeth et al., 2008). Besides array-based gene expression analysis, a number of studies have reported genomic alterations that occur in TNBC clinical specimens and cell lines including comparative genomic hybridization and deep genomic profiling using next generation sequencing (NGS) technologies (Shah et al., 2012; Ding et al., 2010; Weigman et al., 2012; Loo et al., 2011; Neve et al., 2006; Fridlyand et al., 2006; Chin et al., 2006). Whole-genome sequencing of a single metastatic TNBC (mTNBC) patient's germline, primary tumor, metastatic tumor, and xenograft has also been reported, which showed the complexity of the somatic events that arise within a given TNBC (Ding et al., 2010). More recently, genome-sequencing studies of large subsets of retrospectively collected TNBC have been reported, which implicate *TP53*, *PIK3CA*, *NRAS*, *EGFR*, *RB1*, and *PTEN* (Shah et al., 2012; Cancer, 2012).

To identify molecular mechanisms inherent to the TN subtype we have conducted gene set enrichment analysis (GSEA) (Subramanian et al., 2005), comparing TN vs. ER + HER2 −, in seven distinct cohorts, grouping gene sets by common promoter motifs to identify transcription factors and expression patterns of interest. The gene sets that are shown to be enriched in seven distinct cohorts with a Stouffer weighted Z (Whitlock, 2005; Zaykin, 2011) p-value < .01 are used to construct a promoter motif signature for genes determined to be enriched in the maximum number of cohorts. The transcription factor for each identified enriched promoter motif as well as any chemical or genetic perturbation that lowers the expression of the promoter motif gene signature represents potential therapeutic option(s) in TN breast cancer. The workflow is outlined in Fig. 1.

Methods

Cohorts

Cohorts with representation of large N samples with immunohistochemistry (IHC) determined ER +/− and HER2 status and clinical outcome data were selected for analysis. All probe or gene expression levels were used as deposited using published normalization, and the following is a summary of each cohort. Each cohort is molecularly profiled on a wide range of platforms with different normalization methodology. GSEA is done independently for each cohort to determine statistically enriched gene sets mitigating the effects of different platforms and normalizations. The GEO deposited cohorts GSE25055 (n = 279 TN = 114/ER + 165) and GSE25065 (n = 187 TN = 64/123) were run on the U133A Affymetrix GeneChip with well-curated phenotype metadata and metastasis outcome (Hatzis et al., 2011). TCGA-BC RNA Seq V2 RSEM was downloaded from TCGA Data Portal on July 1, 2013 and represents (n = 286 TN = 58/ER + = 228) samples