Orginal Research Article

IN SILICO IDENTIFICATION OF TARGET PALINDROMIC GENES AS POTENTIAL DRUG TARGETS IN BREAST CANCER THERAPY

ABSTRACT

Breast cancer (BC) is the most common cancer worldwide prevalent among women with more than one million cases and is second only to lung cancer. The identification of the sequences based on the unique tetramers GCAC, GTCA were selected from experimental work. The16 base pair DNA regulatory sequences of which the motifs area part of containing these motif in genes impliocated in cancer CAGE1(AAGCTGTCATTA),BRCA1(GACTGAGTCAA),ABCB1(CTCTAAGTCAT),ABCB5 (GATATGTTAAAGC) and ABI1(CTTCTGGGAA) were then selected as novel putative targets in breast cancer therapy based on their selectivity on the BC oncogenes which are not found in the normal human genome 1-23 and the sex chromosomes x and Y.

were obtained via computational analysis.possible the *in silico* discovery of putative anti breast cancer targets of importance in the genome

Key words

Breast Cancer, CAGE 1,BRCA, ABCB , ABI!, Target genes

INTRODUCTION

BC is the most common malignancy in women accounting for about 18% of female cancers worldwide and over half a million diagnosed each year. Its incidence increases with age. It was the second most common cancer in the world with over 1million new cases. In Nigeria, cancer cases have increased by 21% out of which 10% are BC(Jemal, Siegel et al. 2008) .

It originates in the cells of the breast and several genes bearing high-penetrance mutations have been implicated in inherited disposition to BC with BRCA1 and BRCA2 being the most important, BC susceptibility

genes other than BRCA1 AND BRCA2 have been identified. These genes are in two categories:

1) Genes with rare-moderate penetrance (CHEK2,ATM and BRIP) (The CHECK2 Breast Cancer Case Control Consortium, 2004).
2) Genes with low penetrance allele(Consortium 2006).

ABCB1 MDR1 gene encodes a well characterized trans membrane transporter P-glycoprotein. It is expressed in cancer cells and involved in genomic instability of cancer cells, gene rearrangements and mutations. Permeability glycoprotein also known P-glycoprotein (P-gp; *MDR1*; ABCB1) is an efflux transporter which belongs to the ATP-bindingcassette (ABC) superfamily of transporters, and actively transport a wide range of structurally andmechanistically diverse endogenous and xenobiotic chemical agents across the cell membrane atthe energy expense of ATP hydrolysis (Chen et al., 2018). P-glycoprotein has been implicated in the development of multidrug resistance to anticancer drugs when expressed in breast cancer cells (Clarke, Leonessa et al. 2005) due to its ability to profoundly implicate the role of drug pharmacokinetics (PK) that can clinically alter the administrated drugefficacy or even lead to various adverse side-effects due to drug–drug interactions (DDIs) in the caseof polypharmacy (Selick et al., 2002; Montanari and Ecker, 2015; Chen et al., 2018). Presence of oestrogen have been found to downregulate P-glcoprotein expression in ER positive BC cell lines (Imai, Ishikawa et al. 2005).

CAGE1 has been identified and characterized as a novel cancer antigene (Park, Lim et al. 2003).

5-10% of BC cases develop on a hereditary basis. Approximately 80% are accounted for by mutations in the BC susceptibility genes,BRCA1((40-45%) and BRCA2(35-40%).The remaining 20% of BCs may be due to mutations in other tumor suppressor genes as P53,PTEN and ATM in addition to other genes. Genetic events involved in BC tumor formation are as follows: tumor oncogene activation,
Tumor suppressor gene inactivation, abnormal protein over expression and multiple gene alteration(Hudson, Anderson et al. 2010) .

Systemic treatment options include cytotoxic, hormonal and immunotherapeutic agents (Burton, Clayton et al. 2007).In addition, surgery, radiotherapy, hormonal therapy, biological therapy as well as chemotherapy. The chemotherapeutic interventions are used in adjuvant, neoadjuvant, and metastatic settings. In general, systemic agents are active at the beginning of the therapy in 90% of primary BC cases and 50% of metastatic settings. However, after a variable period of time, progression occurs. Resistance to therapy occurs at this point. Most of the existing drugs are toxic, non specific and have severe side effects (Hortobagyi 1998), many tumor develop resistance to majority of the chemotherapeutic agents.
Estrogen receptors ERs are dependent, inducible signal transducers which act through a nonclassical pathway in which liganded ERs are tethered to DNA via association with other transcription factor complexes including Fos/Jun [activator protein-1 (AP-1)–responsive elements] or SP-1(Kushner, Agard et al. 2000, Saville, Wormke et al. 2000) They are capable of binding to palindromic DNA sequences 5′-TGAGTCA-3′ as well as similar sequences 5′-TTAGTCA-3′ and
5′-TGATTCA-3′(Lee, Hahn et al. 2000, Papassava, Gorgoulis et al. 2004)
A small number of genes containing an AP-1 site in their promoters have been shown to be regulated by ERs(Schmitt, Bausero et al. 1995) . AP-1 can be induced by extracellular stimuli such as cytokines, UV radiation, growth factors, oxidative stress, and carcinogens can and promote AP-1 binding to the TRE of its target genes that are involved in cell growth, inflammatory responses, and repair processes.
In cellular and viral genes, the AP-1 transcription factor is thus a major component of many signal transduction pathways. It is a complex dimer of homo- and heterodimer family members. The regulation of AP-1 target genes is regulated by AP-1 transcription factor via binding to the DNA regulatory sequence 12. The response TPA element sequence GCAC or GTCA are therefore bound by c-Jun homodimers and c-Jun/c-Fos heterodimers (Van Dam and Castellazzi 2001).
In this research work, DNA regulatory sequence motifs containing the unique tetramer GCAC, GTCA have been curated from experimental work and
The post genomic era characterized by the completion of the human genome project, the start of both the proteomics and structural genomics revolutions and the development in information technology have resulted in the use of structure based drug design in the discovery of new drug targets. The human genome is the complete set of genetic information for humans (*Homo sapiens)*. This information is encoded as DNA sequences within the 23 chromosome pairs (22 pairs being homologous) in cell nuclei and in a small DNA molecule found in the mitochondria. It includes both the protein-coding genes as well as the non-coding genes. The haploid human genome (contained in the spermatozoa and oocyte) consist of three billion DNA base pairs while the diploid genomes found in somatic cells have twice the c content (Guerzoni and McLysaght 2011) There is an urgent need to identify novel putative targets into the breast cancer drug development pipeline as an early intervention method in breast cancer therapy . *In silico* intervention at various stages in the drug discovery cycle is equipped with the potentials to reduce both time and cost involved in the process (Collins, Patrinos et al. 1998) The interaction between small molecules and DNA is important due to its implication in the regulation of gene expression by activators and repressors in vivo (Paleček, Fojta et al. 1998)
DNA as a drug target in the case of infectious agents and that of the human as in BC have proved attractive due to the availability of the three dimensional DNA structures and the

predictability of their accessible chemical functional

In view of the availability of such novel targets which to which selective chemotherapeutic agents against BC can be developed, 40% still die ultimately from the disease. These warrant the need for the development of novel therapeutic agents in the treatment of BC via identification of potential drug targets.

Method

The entire Human genome was downloaded via the file transfer protocol ftp from the database of the National Centre for Biotechnology Information (NCBI) ftp site . Homo_Sapiens.NCBI34.pep.fa.downloaded from ftpsite (ftp.// ftp.ensembl.org/pub/current_human/data/fasta/pep/). The download was done in the fasta format. The oncogenes BRCA, CAGE1, ABCB5 and ABI1 were also downloaded from the Online Mendelian Inheritance In Man was also downloaded from the website where OMIM is domiciled.

https://bioportal.bioontology.org/ontologies/OMIM . The selectivity of the oncogenes as potential target agents was ascertained when the sequence frequency of occurrence in the human genome was determined.

Results

Table 1.0 Summary of selected oncogenes, start/stop positions on the specific chromosome, gene sequence and the base pair number

Gene Name	Length	Transcription Factor	Chromosome Number	Binding Position	Base pairs	BP no
BRCA1	26	AP-1	Chr7	41267898-41267908-	GACTGAGTCAA	11
CAGE 1	3	TGIF	Chr6	7405451-7405462	AAGCTGTCATTA	12
ABCB1		AP-1	Chr7	87336139-87336149	CTCTAAGTCAT	11

Table 2.0 16base pair sequences as potential drug targets from 11bp BRCA1

Gene Name			
BRCA1	GACTGA**GTCA**A	11	615
	TGACTGAGTCAA	12	197
	ATGACTGAGTCAA	13	75
	TATGACTGAGTCAA	14	14
	TTATGACTGAGTCAA	15	4
	GTTATGACTGAGTCAA	16	2
	TGTTATGACTGAGTCAA	17	2
	CTGTTATGACTGAGTCAA	18	1

Table 3.0 16base pair sequences as potential drug targets from 11base pair sequence of CAGE1

CAGE1	AAGCT**GTCA**TTA	12	328

	AAAGCTGTCATTA	13	122
	AAAAGCTGTCATTA	14	37
	TAAAAGCTGTCATTA	15	9
	ATAAAAGCTGTCATTA	16	5
	CATAAAAGCTGTCATTA	17	1

Table 4.0 16base pair sequences as potential drug targets from 11base pair of ABCB1 oncogene

ABCB1	CTCTAA**GTCA**T	11	810
	ACTCTAAGTCAT	12	199
	AACTCTAAGTCAT	13	63
	CAACTCTAAGTCAT	14	13
	CCAACTCTAAGTCAT	15	6
	GCCAACTCTAAGTCAT	16	3
	TGCCAACTCTAAGTCAT	17	2
	TTGCCAACTCTAAGTCAT	18	1

Discussion

The selected genes have been ABCB1, CAGE1 and BRCA1 have been implicated in the proliferation and development of cancer cells. The base pairs numbers of these genes are as follows 11 badse pairs for BRCA1 and ABCB1 while CAGE1 has 12 base pairs Table 1.0. The sequences contain GCAC and GTCA which are unique tetramers however the frequency of occurrence of the sequences 615 for BRCA1 Table 2.0 ABCB1 11 base pair CTCTAA**GTCA**T occurs at a frequency of 810 in the human genome while CAGE1 with 12 base pairs AAGCT**GTCA**TTA occurs 328 times. The criteria for an ideal target is that it must occur in a frequency of 1 for it to be non toxic that is it must be a single copy gene. Therefore to obtain a single copy gene from these oncogenes, the start and stop positions wereidentified such that the the sequences to the right and left of the core sequences were determined Table 2.0,3.0 and 4.0. Thus stepwise, the final 16base pairs. The single copy base pairs which will be potential drug targets as anticancer drugs were finally obtained as CTGTTATGACTGAGTCAA, CAGE1 with the 17base pairs CATAAAAGCTGTCATTA and ABCB1 TTGCCAACTCTAAGTCAT .

Conclusion

Potential anti breast cancer targets which will go a long way when drugs are developed against them include

CTGTTATGACTGAGTCAA, CAGE1 with the 17base pairs CATAAAAGCTGTCATTA and ABCB1 TTGCCAACTCTAAGTCAT .Further research work based on drug design techniques will be a be developed.

References-

Burton, P. R., et al. (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature 447(7145): 661-678.

Chen C, Lee MH, Weng CF, and Leong MK, (2018). Theoretical Prediction of the Complex P-Glycoprotein Substrate Efflux Based on the Novel Hierarchical Support Vector Regression Scheme.Molecules, 23, 1820; doi:10.3390/molecules23071820

Clarke, R., et al. (2005). Multidrug resistance/P-glycoprotein and breast cancer: review and meta-analysis. Seminars in Oncology, Elsevier.

Collins, F. S., et al. (1998). "New goals for the US human genome project: 1998-2003." Science 282(5389): 682-689.

Consortium, B. C. A. (2006). "Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium." Journal of the National Cancer Institute 98(19): 1382-1396.

Guerzoni, D. and A. McLysaght (2011). "De novo origins of human genes." PLoS genetics 7(11): e1002381.

Hortobagyi, G. N. (1998). "Treatment of breast cancer." New England Journal of Medicine 339(14): 974-984.

Hudson, T. J., et al. (2010). "International network of cancer genome projects." Nature 464(7291): 993-998.

Imai, Y., et al. (2005). "Estrogen-mediated post transcriptional down-regulation of breast cancer resistance protein/ABCG2." Cancer Research 65(2): 596-604.

Jemal, A., et al. (2008). "Cancer statistics, 2008." CA: A Cancer Journal for Clinicians 58(2): 71-96.

Kushner, P. J., et al. (2000). "Estrogen receptor pathways to AP-1." The Journal of steroid biochemistry and molecular biology 74(5): 311-317.

Lee, J., et al. (2000). "Characterization of JDP genes, an evolutionarily conserved J domain-only protein family, from human and moths." Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1491(1): 355-363.

Paleček, E., et al. (1998). "Electrochemical biosensors for DNA hybridization and DNA damage." Biosensors and Bioelectronics 13(6): 621-628.

Papassava, P., et al. (2004). "Overexpression of activating transcription factor-2 is required for tumor growth and progression in mouse skin tumors." Cancer Research 64(23): 8573-8584.

Park, S., et al. (2003). "Identification and characterization of a novel cancer/testis antigen gene CAGE-1." Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1625(2): 173-182.

Saville, B., et al. (2000). "Ligand-, cell-, and estrogen receptor subtype (α/β)-dependent activation at GC-rich (Sp1) promoter elements." Journal of Biological Chemistry 275(8): 5379-5387.

Schmitt, M., et al. (1995). "Positive and negative effects of nuclear receptors on transcription activation by AP-1 of the human choline acetyltransferase proximal promoter." Journal of Neuroscience Research 40(2): 152-164.

Van Dam, H. and M. Castellazzi (2001). "Distinct roles of Jun: Fos and Jun: ATF dimers in oncogenesis." Oncogene 20(19): 2453.