Advanced Techniques in Biology & Medicine
ISSN: 2379-1764

Yadav and Jha, Adv Tech Biol Med 2015, 3:3
DOI: 10.4172/2379-1764.1000147

Research Article Open Access

Metagenomic Analysis of Molecular Profile of Breast Cancer Using Genie a Literature Based Gene Prioritizing Tool: A Novel Approach

Amit Kumar Yadav* and Vidya Jha

Division of Molecular Pathology, Department of Pathology, Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi, India

Abstract

Background: The biological complexity and heterogeneity of breast cancer can be explained by molecular profile. Microarray technique is the method of choice to study it. But there are certain limitations. Computational techniques are a novel method to study gene expression profile.

Materials and method: Genie, a freely available web based software was used to analyze literature, gene and homology information from MEDLINE, NCBI Gene and HomoloGene databases. Inputs given were target species (Homo sapiens) and biomedical topic (breast cancer). According to input provided genes of target species are prioritized.

Results: The ranking given to 1906 reported genes was not along expected lines. Therefore, they were manually re-ranked according to number of hits. These were narrowed to 70. The proteins encoded by these genes and their functions were obtained from NCBI database. On the basis of their function and role in carcinogenesis these genes were then grouped together into distinct categories.

Conclusion: A novel computational approach to study molecular profile of breast cancer has been demonstrated. A panel of 70 genes to study gene expression profile of breast cancer has been suggested. These genes comprehensively evaluate all aspects of molecular pathogenesis of breast cancer and recommended for future clinical studies.

Keywords: Breast cancer; Molecular profile; Genie; Gene prioritization; Metagenomic analysis

Introduction

Breast cancer is a very complex and heterogeneous disease. The vast majority of cases are morphologically infiltrating ductal carcinoma NOS. However, in these morphologically similar looking cases the biological behaviour is different. This leads to difference in response to therapy and outcome. Clinical parameters like age, tumour stage, grade and routinely used biomarkers like estrogen receptor (ER), progesterone receptor (PR) and HER2-neu cannot fully explain this heterogeneity [1,2]. This is due to the difference in molecular profile of each case.

Traditional classification of breast cancer is based on morphology. The study of gene expression profile has led to a new system of classification. In 2000, Perou and co-workers [3] in a seminal paper classified breast cancer into intrinsic subtypes based on gene expression profile. Subsequent work of numerous authors has fundamentally changed the way breast cancer is understood and classified [4-6].

Microarray technique which allows simultaneous analysis of expression of thousands of genes has been the method of choice to study gene expression profile of breast cancer [7]. The disadvantage of this technique is the limited sample and variability [8,9]. Also a comprehensive summarization of the genes is not possible due to the large number of genes and abstracts involved.

A novel approach to study molecular profile is using computational techniques to study gene function by analyzing the vast literature which is now available. Fontaine and co-workers [10] developed the Genie algorithm and web server. The input for the software is a biological topic.

It evaluates the entire MEDLINE database for relevance to that subject, and then evaluates all the genes of a user's requested organism according to the relevance of their associated MEDLINE records. The advantage of this approach is that large amount of published data regarding genes in a particular disease can be analyzed. This kind of analysis at a multi genomic scale is not possible without computational approach. To the best of our knowledge this approach has not been used to date for studying molecular profile of breast cancer.

Materials and Method

Genie is freely available to all researchers at http://cbdm.mdc-berlin.de/tools/genie/. The algorithm and web server takes advantage of literature, gene and homology information from the MEDLINE, NCBI Gene and HomoloGene databases [8].

The system requires two basic inputs: a target species (e.g., *Homo sapiens*) and a biomedical topic ideally related to a gene function (breast cancer in the present study). According to the input provided the genes of the target species are prioritized. The target species is defined by its scientific name or its taxonomic ID (e.g., *Homo sapiens* - 9606).

The biomedical topic is ultimately defined by a set of biomedical references represented by MEDLINE records. After giving the inputs the software initialised in 9 seconds and the analysis was complete in 73 seconds. It went through 796 abstracts on PubMed. All the relevant protein coding genes as per the input provided were analyzed. In order to ensure that only significant genes were reported the cut offs were

***Corresponding author:** Amit Kumar Yadav, Assistant Professor, Division of Molecular Pathology, Department of Pathology, Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi, India, Tel: 091-26707408; E-mail: amityadav7284@yahoo.co.in

Received October 14, 2015; **Accepted** October 24, 2015; **Published** October 31, 2015

Citation: Yadav AK, Jha V (2015) Metagenomic Analysis of Molecular Profile of Breast Cancer Using Genie a Literature Based Gene Prioritizing Tool: A Novel Approach. Adv Tech Biol Med 3: 147. doi: 10.4172/2379-1764.1000147

taken as p<0.01 for abstracts and false discovery rate<0.01 for genes. A one-sided Fisher's exact test was carried out by the algorithm to define the significance of gene-to topic relationship. It compared the number of selected abstracts to what is observed in a simulation using a set of ten thousand randomly selected abstracts.

Literature extension by orthology was not done as it was not needed. This is done when genes from poorly studied organisms are studied. The genes are ranked using, in addition to the abstracts directly associated to them, the abstracts associated to their orthologs in other species.

The genes are then presented in a list sorted by false discovery rate (FDR) with hyperlinks to the most significant abstracts, Entrez Gene and HomoloGene databases. A list of the words found to be relevant to the topic is provided to facilitate the interpretation of the results.

Results

A total number of 1906 genes were reported by the software to be associated with breast cancer. These genes were prioritized and ranked according to the abstracts directly associated with them. It was observed that the ranking given by the software was not along expected lines as it did not correlate with the data available from previous work [11-14]. For example, ZNF703 (zinc finger protein 703) was given first rank while ERBB2 (erb-b2 receptor tyrosine kinase) and ESR1 (estrogen receptor 1) were ranked third and seventh respectively.

In order to remove this discrepancy an alternative method of ranking was devised. The genes were ranked manually according to the number of hits that is the gene with the maximum number of hits was ranked first and one with minimum number of hits was ranked last. In order to find out the most important genes different cut offs for the number of hits were used – 50 and 100 (Table 1). The number of cut offs was arrived at by trial method.

When cut off was used as 50 a total of 70 genes were selected, whereas a cut off of 100 yielded 33 genes. The first 5 ranks are now occupied by ESR1 (estrogen receptor 1), ERBB2 (erb-b2 receptor tyrosine kinase 2), EGFR (epidermal growth factor receptor), TP53 (tumor protein p53) and BRCA1 (breast cancer 1, early onset) genes. Previously these were taken by ZNF703 (zinc finger protein 703), GREB1 (Growth regulation by estrogen in breast cancer 1), ERBB2 (erb-b2 receptor tyrosine kinase 2), CST6 (cystatin E/M) and WISP2 (WNT1 inducible signalling pathway protein 2).

Rank	Gene ID	Symbol	PMID	Hits
1	2099	ESR1	2339	851
2	2064	ERBB2	1756	733
3	1956	EGFR	3346	492
4	7157	TP53	6528	483
5	672	BRCA1	2027	405
6	207	AKT1	1997	263
7	2100	ESR2	875	255
8	7422	VEGFA	3104	243
9	3091	HIF1A	1901	188
10	595	CCND1	1116	178
11	5241	PGR	555	177
12	5728	PTEN	1298	177
13	6774	STAT3	1459	169
14	5743	PTGS2	1745	169
15	367	AR	1634	159
16	999	CDH1	1306	155
17	5290	PIK3CA	807	153
18	7040	TGFB1	2828	150
19	675	BRCA2	1223	148
20	3480	IGF1R	690	141
21	596	BCL2	1374	142
22	4318	MMP9	1862	141
23	7852	CXCR4	1335	137
24	4790	NFKB1	2235	134
25	332	BIRC5	888	134
26	1499	CTNNB1	1609	131
27	1026	CDKN1A	1202	121
28	1588	CYP19A1	596	111
29	4233	MET	660	111
30	7316	UBC	3431	111
31	960	CD44	751	106
32	6714	SRC	1032	104
33	1027	CDKN1B	762	100
34	3845	KRAS	1280	99
35	5594	MAPK1	1555	99
36	2475	MTOR	981	97
37	4609	MYC	1180	94
38	8202	NCOA3	278	89
39	3479	IGF1	1278	84
40	1029	CDKN2A	1654	82
41	5595	MAPK3	1074	81
42	4313	MMP2	1208	80
43	4582	MUC1	524	78
44	9429	ABCG2	517	76
45	3569	IL6	3074	76
46	7424	VEGFC	308	75
47	2146	EZH2	406	75
48	5243	ABCB1	1590	74
49	4288	MKI67	449	73
50	6387	CXCL12	885	70
51	5468	PPARG	1563	69
52	2065	ERBB3	289	69
53	8743	TNFSF10	606	64
54	6696	SPP1	657	64
55	4851	NOTCH1	714	63
56	5747	PTK2	607	61
57	6667	SP1	833	59
58	7124	TNF	4360	56
59	3065	HDAC1	803	56
60	857	CAV1	637	55
61	5970	RELA	976	55
62	5268	SERPINB5	167	54
63	100133941	CD24	180	54
64	6615	SNAI1	263	54
65	5328	PLAU	463	54
66	11186	RASSF1	334	53
67	4193	MDM2	1205	53
68	5925	RB1	899	51
69	3486	IGFBP3	535	50
70	2950	GSTP1	1248	50

Table 1: Genes reported by genie software with their rank, gene ID, symbol, PMID and number of hits.

The proteins encoded by these genes and their functions were obtained from NCBI database. These are summarized in Table 2. On

Rank	Gene	Protein encoded	Function
1	**ESR1**	Estrogen receptor 1	Localizes to nucleus and binds estrogen hormone.
2	**ERBB2**	erb-b2 receptor tyrosine kinase 2	A member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases. It forms a heterodimer, stabilizing ligand binding and enhancing kinase-mediated activation of downstream signalling pathways,
3	**EGFR**	Epidermal growth factor receptor	A transmembrane glycoprotein which is a receptor for members of the epidermal growth factor family. Binding of the protein to ligand leads to cell proliferation.
4	**TP53**	Tumour protein p53	Tumour suppressor protein
5	**BRCA1**	Breast cancer 1, early onset	Nuclear phosphoprotein playing role in maintaining genomic stability and acting as a tumor suppressor.
6	**AKT1**	V-akt murine thymoma viral oncogene homolog 1	Role in cell survival and inhibition of apoptosis.
7	**ESR2**	Estrogen receptor 2	Binds to estrogen hormone and interact with specific DNA sequences to activate transcription.
8	**VEGFA**	Vascular endothelial growth factor a	Acts on endothelial cells and leads to increased vascular permeability, inducing angiogenesis, vasculogenesis and endothelial cell growth, promoting cell migration, and inhibiting apoptosis.
9	**HIF1A**	Hypoxia inducible factor 1, alpha subunit	Master regulator of cellular and systemic homeostatic response to hypoxia by activating transcription of many genes.
10	**CCND1**	Cyclin d1	Regulation of cell cycle.
11	**PGR**	Progesterone receptor	Mediates action of progesterone hormone.
12	**PTEN**	Phosphatase and tensin homolog	Tumour suppressor protein
13	**STAT3**	Signal transducer and activator of transcription 3	Act as transcriptional activators to many genes.
14	**PTGS2**	Prostaglandin-endoperoxide synthase 2	Prostaglandin biosynthesis
15	**AR**	Androgen receptor.	Stimulates transcription of androgen responsive genes.
16	**CDH1**	Cadherin 1, type 1, e-cadherin (epithelial)	Mediates cell-cell adhesion.
17	**PIK3CA**	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	Catalytic subunit of Phosphatidylinositol 3-kinase
18	**TGFB1**	Transforming growth factor, beta 1	Regulates proliferation, differentiation, adhesion, migration and other functions.
19	**BRCA2**	Breast cancer 2, early onset	Maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair.
20	**IGF1R**	Insulin-like growth factor 1 receptor	Binds insulin-like growth factor and functions as an anti-apoptotic agent by enhancing cell survival.
21	**BCL2**	B-cell cll/lymphoma 2	Integral outer mitochondrial membrane protein that blocks apoptosis.
22	**MMP9**	Matrix metallopeptidase 9	Degrades type IV and V collagens.
23	**CXCR4**	Chemokine (c-x-c motif) receptor 4	Encodes a CXC chemokine receptor specific for stromal cell-derived factor-1.
24	**NFKB1**	Nuclear factor of kappa light polypeptide gene enhancer in b-cells 1	DNA binding subunit of the NF-kappa-B (NFKB) protein complex.
25	**BIRC5**	Baculoviral iap repeat containing 5	Encode negative regulatory proteins that prevent apoptotic cell death.
26	**CTNNB1**	Catenin (cadherin-associated protein), beta 1, 88kda	Part of a complex of proteins that constitute adherens junctions (AJs) and may be responsible for transmitting the contact inhibition signal that causes cells to stop dividing once the epithelial sheet is complete.
27	**CDKN1A**	Cyclin-dependent kinase inhibitor 1a (p21, cip1)	Regulator of cell cycle progression at G1.
28	**CYP19A1**	Cytochrome p450, family 19, subfamily a, polypeptide 1	Member of the cytochrome P450 superfamily of enzymes. This protein localizes to the endoplasmic reticulum and catalyses the last steps of estrogen biosynthesis.
29	**MET**	Met proto-oncogene, receptor tyrosine kinase	Encodes tyrosine-kinase activity.
30	**UBC**	Ubiquitin c	Protein degradation, DNA repairs, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signalling pathways.
31	**CD44**	Cd44 molecule (Indian blood group)	Participates in a wide variety of cellular functions including lymphocyte activation, recirculation and homing, hematopoiesis, and tumor metastasis.
32	**SRC**	Src proto-oncogene, non-receptor tyrosine kinase	Regulation of embryonic development and cell growth.
33	**CDKN1B**	Cyclin-dependent kinase inhibitor 1b	Controls the cell cycle progression at G1.
34	**KRAS**	Kirsten rat sarcoma viral oncogene homolog	Member of the small GTPase superfamily.
35	**MAPK1**	Mitogen-activated protein kinase 1	Integration point for multiple biochemical signals, and involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development.
36	**MTOR**	Mechanistic target of rapamycin (serine/threonine kinase)	Mediate cellular responses to stresses such as DNA damage and nutrient deprivation.
37	**MYC**	V-myc avian myelocytomatosis viral oncogene homolog	Role in cell cycle progression, apoptosis and cellular transformation.
38	**NCOA3**	Nuclear receptor coactivator 3	Nuclear receptor coactivator that interacts with nuclear hormone receptors to enhance their transcriptional activator functions.
39	**IGF1**	Insulin-like growth factor 1	Member of a family of proteins involved in mediating growth and development.
40	**CDKN2A**	Cyclin-dependent kinase inhibitor 2a	Involved in cell cycle G1 control and is an important tumor suppressor gene.
41	**MAPK3**	Mitogen-activated protein kinase 3	Act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals.
42	**MMP2**	Matrix metallopeptidase 2	Zinc-dependent enzymes capable of cleaving components of the extracellular matrix and molecules involved in signal transduction.

43	**MUC1**	Mucin 1, cell surface associated	Membrane-bound protein playing an essential role in forming protective mucous barriers on epithelial surfaces and intracellular signaling.
44	**ABCG2**	Atp-binding cassette, sub-family g (white), member 2 (junior blood group)	Transport various molecules across extra- and intra-cellular membranes. Also referred to as a breast cancer resistance protein, this protein functions as a xenobiotic transporter which may play a major role in multi-drug resistance.
45	**IL6**	Interleukin 6	Cytokine that functions in inflammation and the maturation of B cells.
46	**VEGFC**	Vascular endothelial growth factor c	Promotes angiogenesis and endothelial cell growth, and can also affect the permeability of blood vessels.
47	**EZH2**	Enhancer of zeste 2 polycomb repressive complex 2 subunit	Involved in maintaining the transcriptional repressive state of genes over successive cell generations.
48	**ABCB1**	Atp-binding cassette, sub-family b (mdr/tap), member 1	Transport various molecules across extra- and intra-cellular membranes. Involved in multidrug resistance by acting as an ATP-dependent drug efflux pump. It often mediates the development of resistance to anticancer drugs.
49	**MKI67**	Marker of proliferation ki-67	Associated with and may be necessary for cellular proliferation.
50	**CXCL12**	Chemokine (c-x-c motif) ligand 12	Functions as the ligand for the G-protein coupled receptor, chemokine (C-X-C motif) receptor 4, and plays a role in many diverse cellular functions, including embryogenesis, immune surveillance, inflammation response, tissue homeostasis, and tumor growth and metastasis.
51	**PPARG**	Peroxisome proliferator-activated receptor gamma	Regulator of adipocyte differentiation.
52	**ERBB3**	Erb-b2 receptor tyrosine kinase 3	Forms heterodimers with other EGF receptor family members having kinase activity. This leads to the activation of pathways which lead to cell proliferation or differentiation.
53	**TNFSF10**	Tumor necrosis factor (ligand) superfamily, member 10	Preferentially induces apoptosis in transformed and tumor cells, but does not appear to kill normal cells although it is expressed at a significant level in most normal tissues.
54	**SPP1**	Secreted phosphoprotein 1	Cytokine that upregulates expression of interferon-gamma and interleukin-12.
55	**NOTCH1**	Notch 1	Play a role in a variety of developmental processes by controlling cell fate decisions.
56	**PTK2**	Protein tyrosine kinase 2	Cell growth and intracellular signal transduction pathways triggered in response to certain neural peptides or to cell interactions with the extracellular matrix.
57	**SP1**	Sp1 transcription factor	Zinc finger transcription factor involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling.
58	**TNF**	Tumor necrosis factor	Regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation.
59	**HDAC1**	Histone deacetylase 1	Component of the histone deacetylase complex. It also interacts with retinoblastoma tumor-suppressor protein and this complex is a key element in the control of cell proliferation and differentiation.
60	**CAV1**	Caveolin 1, caveolae protein, 22kda	Promote cell cycle progression and tumor suppressor gene
61	**RELA**	V-rel avian reticuloendotheliosis viral oncogene homolog a	Inhibitor of NF-kappa-B (NFKB1).
62	**SERPINB5**	Serpin peptidase inhibitor, clade b (ovalbumin), member 5	Tumor suppressor. It blocks the growth, invasion, and metastatic properties of mammary tumors.
63	**CD24**	CD24 molecule	Modulates growth and differentiation signals to granulocytes and B cells.
64	**SNAI1**	Snail family zinc finger 1	Zinc finger transcriptional repressor which downregulates the expression of ectodermal genes within the mesoderm.
65	**PLAU**	Plasminogen activator, urokinase	Serine protease involved in degradation of the extracellular matrix and possibly tumor cell migration and proliferation.
66	**RASSF1**	Ras association (ralgds/af-6) domain family member 1	Tumour suppressor function.
67	**MDM2**	Mdm2 proto-oncogene, e3 ubiquitin protein ligase	Promote tumor formation by targeting tumor suppressor proteins, such as p53, for proteasomal degradation.
68	**RB1**	Retinoblastoma 1	Negative regulator of the cell cycle and a tumor suppressor gene.
69	**IGFBP3**	Insulin-like growth factor binding protein 3	Protein forms a ternary complex with insulin-like growth factor acid-labile subunit (IGFALS) and either insulin-like growth factor (IGF) I or II. It circulates in the plasma, prolonging the half-life of IGFs and altering their interaction with cell surface receptors.
70	**GSTP1**	Glutathione s-transferase pi 1	Function in xenobiotic metabolism and play a role in susceptibility to cancer, and other diseases.

Table 2: Proteins encoded by the genes and their function.

the basis of their function and role in carcinogenesis these genes were then grouped together into following distinct categories (Figure 1)

a) Proliferation.

b) Evading apoptosis.

c) Invasion and metastasis.

d) Sustained angiogenesis.

e) Tumour suppressor genes.

f) Estrogen.

g) Her-2 neu.

h) Miscellaneous.

Discussion

Breast cancer is a leading cause of cancer related mortality in women. As per WHO statistics [15] nearly 1.7 million new cases were diagnosed in 2012 (second most common cancer overall). This represents about 12% of all new cancer cases and 25% of all cancers in women. Traditionally clinical parameters like age, tumour stage, grade and routinely used biomarkers like oestrogen receptor(ER), progesterone receptor (PR) and HER2-neu have been used to evaluate the prognosis and guide therapy [16,17].

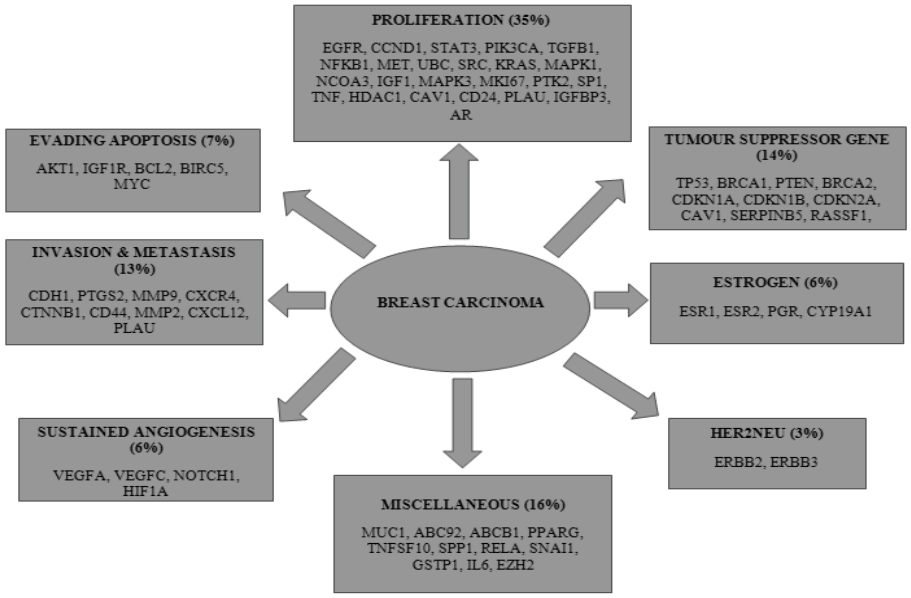

Figure 1: Categorization of genes according to their function and role in genesis of breast cancer.

Vast amount of molecular information in breast cancer is now available from gene expression profiling studies. These have become an extremely important tool to assess the prognosis and guide appropriate management [18]. The technique of choice for this is microarray. The data that has become available from gene expression profile studies has impacted not only the management of breast cancer but other tumours like lung [19,20] and colon cancer [21]. However, microarray technique suffers from many disadvantages as mentioned previously [8,9]. Thus there is a need to explore alternative methods to study gene expression profile of breast cancer.

The authors have used one such technique Genie algorithm and web server in the present study to evaluate molecular profile of breast cancer. The utility of computational approach to find human genes associated with a disease has been shown previously [22-24]. However, Genie goes one step ahead and besides highlighting well known genes it brings out new candidate genes [10]. This helps in better characterization of a disease.

There are alternative gene prioritizing tools that perform automatic gene name extraction and normalization [25,26]. The basic concept behind such an analysis is that when two words repeatedly occur together in an abstract they are likely to be functionally related. However, these methods suffer from a disadvantage that they wrongly identify genes in text which leads to ambiguous results [27,28]. Genie overcomes this limitation by using NCBI curated gene associations and unambiguous gene identifiers [10].

A total number of 1906 genes were found to be associated with breast cancer. These genes were manually ranked using the number of hits as a parameter. Then the number of genes was further narrowed down by using two cut offs, i.e., 50 and 100 hits. These yielded 70 and 33 genes respectively. The authors feel that 33 genes are not sufficient for a comprehensive analysis of gene expression profile. Thus it is recommended that at least 70 genes should be studied for a proper examination. On categorization of these genes on the basis of function it was observed that amongst all the categories maximum number of genes belonged to the proliferation related group (24 genes). This was followed by tumour suppressor genes (12 genes). Overall the genes reported by the software comprehensively cover all aspects of the biology of breast cancer.

The availability of new data from microarray studies led to the development of many multigene prognostic tests to improve assessment of prognosis and therapeutic response in breast cancer [29]. The most widely used amongst these are Oncotype DX [14] and Mammaprint [13].

Oncotype DX (Genomic Health, Redwood City, CA, USA) is based on high-throughput real time, reverse transcriptase polymerase chain reaction (RT-PCR) analysis of formalin fixed paraffin-embedded (FFPE) tumor tissue [30,31]. Thus, it can also be used on archival blocks. The test utilizes 16 genes which have been shown to have highest correlation with distant recurrence after 10 years along with five housekeeping genes. The test algorithm is designed to calculate recurrence score (RS) from 0 to 100. A higher RS is associated with greater probability of recurrence at 10 years and vice versa.

MammaPrint (Agilent, Amsterdam, Netherlands) is a microarray-based test. It measures the expression of 70 genes. The test is recommended as an adjunctive prognostic test for breast cancer patients who are less than 61 years of age with stage I/II disease, lymph node-negative or one to three lymph node-positive [12]. MammaPrint stratifies patients into low-risk or high-risk prognostic groups [13]. The prognostic risk discrimination is good among. In patients who are ER-positive the assay has a good prognostic risk assessment. However, almost all ER-negative cases are stratified as high risk. This makes the prognostic score of limited clinical value in this group [32].

A large multicenter retrospective study suggested that adjuvant chemotherapy was beneficial only in the high-risk ER positive patients [33]. MammaPrint as described originally needed fresh-frozen tissue. This was a major drawback and reason for its limited clinical utilization. However, recently described version of the test can be performed on FFPE tissue [34].

Using a novel computational technique to carry out molecular

profiling at a multi-genomic scale the authors suggest an alternative panel of genes (Table 1) to study gene expression profile of breast cancer. It is believed that this panel includes some of the important genes which were not present in the initial panels.

A comparison was done between the panel suggested in the present study and those included in Oncotype DX [14] and MammaPrint [13]. In case of Oncotype DX out of the total 16 genes related to breast cancer 6 (38%) are also present in the new panel. While in the case of MammaPrint out of the total 70 only a single gene i.e. MMP9 is shared with our panel. Thus there was greater correlation with Oncotype DX as compared to MammaPrint.

In a recent review of clinical utility of gene-expression profiling in women with early breast cancer Marrone and his co-workers [35] have said that five systematic reviews found no direct evidence of clinical utility for either Oncotype DX or MammaPrint. Indirect evidence showed Oncotype DX was able to predict treatment effects of adjuvant chemotherapy, whereas no evidence of predictive value was found for MammaPrint. No studies provided any direct evidence that using gene-expression profiling tests to direct treatment decisions improved outcomes in women with breast cancer. The authors believe that one of the main reasons for this apparent failure of the two techniques to influence treatment outcome is probably inappropriate gene selection. On going through the list of genes in both the tests it was felt that possibly some important genes have not been included. Thus there is a need for an alternative panel.

Conclusion

The present study has demonstrated a novel computational approach to study molecular profile of breast cancer using genie, a gene prioritizing software. A novel panel of 70 genes to study gene expression profile of breast cancer has been suggested in which the majority of genes are different from currently used panels. These genes we believe comprehensively evaluate all aspects of the molecular pathogenesis of breast cancer. However, the clinical utility of the present study can only be found out by carrying out well designed clinical studies in the future.

References

1. Foulkes WD, Reis-Filho JS, Narod SA (2010) Tumor size and survival in breast cancer--a reappraisal. Nat Rev Clin Oncol 7: 348-353.
2. Foulkes WD, Grainge MJ, Rakha EA, Green AR, Ellis IO (2009) Tumor size is an unreliable predictor of prognosis in basal-like breast cancers and does not correlate closely with lymph node status. Breast Cancer Res Treat 117: 199-204.
3. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. Nature 406: 747-752.
4. Reis-Filho JS, Pusztai L (2011) Gene expression profiling in breast cancer: classification, prognostication, and prediction. Lancet 378: 1812-1823.
5. Hu Z, Fan C, Oh DS, Marron JS, He X, et al. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7: 96.
6. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, et al. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100: 10393-10398.
7. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci 98: 10869-10874.
8. Mackay A, Weigelt B, Grigoriadis A, Kreike B, Natrajan R, et al. (2011) Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. J Natl Cancer Inst 103: 662-673.
9. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, et al. (2006) Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355: 560-569.
10. Fontaine JF, Priller F, Barbosa-Silva A, Andrade-Navarro MA (2011) Génie: literature-based gene prioritization at multi genomic scale. Nucleic Acids Res 39: W455-461.
11. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, et al. (2010) Database resources of the national center for biotechnology information. Nucleic Acids Res 38: 5-16.
12. Mook S, Schmidt MK, Viale G, Pruneri G, Eekhout I, et al. (2009) The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. Breast Cancer Res Treat 116: 295-302.
13. Van de Vijver MJ, He YD, Van't Veer LJ, Dai H, Hart AA, et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347: 1999-2009.
14. Paik S, Shak S, Tang G, Kim C, Baker J, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351: 2817-2826.
15. Ferlay J, Soerjomataram I, Ervik M (2012) GLOBOCAN 2012: Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer.
16. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, et al. (2007) American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 25: 5287-5312.
17. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, et al. (2011) Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol 29: 4273-4278.
18. Sotiriou C, Pusztai L (2009) Gene-expression signatures in breast cancer. N Engl J Med 360: 790-800.
19. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, et al. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350: 2129-2139.
20. Pao W, Miller V, Zakowski M, Doherty J, Politi K, et al. (2004) EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A 101: 13306-13311.
21. Van Cutsem E, Kohne CH, Lang I (2011) Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status. J Clin Oncol 29: 2011–2019.
22. Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37: W305-311.
23. van Driel MA, Cuelenaere K, Kemmeren PP, Leunissen JA, Brunner HG, et al. (2005) GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases. Nucleic Acids Res 33: W758-761.
24. Perez-Iratxeta C, Bork P, Andrade-Navarro MA (2007) Update of the G2D tool for prioritization of gene candidates to inherited diseases. Nucleic Acids Res 35: W212-216.
25. Crim J, McDonald R, Pereira F (2005) Automatically annotating documents with normalized gene lists. BMC Bioinformatics 6 Suppl 1: S13.
26. Cheng D, Knox C, Young N, Stothard P, Damaraju S, et al. (2008) PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. Nucleic Acids Res 36: 399-405.
27. Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, et al. (2008) Overview of BioCreative II gene normalization. Genome Biol 9 Suppl 2: S3.
28. Wermter J, Tomanek K, Hahn U (2009) High-performance gene name normalization with GeNo. Bioinformatics 25: 815-821.
29. GyÅ'rffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, et al. (2015) Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Res 17: 11.
30. Cronin M, Pho M, Dutta D, Stephans JC, Shak S, et al. (2004) Measurement

of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. Am J Pathol 164: 35-42.

31. Cronin M, Sangli C, Liu ML, Pho M, Dutta D, et al. (2007) Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. Clin Chem 53: 1084-1091.

32. Knauer M, Mook S, Rutgers EJ, Bender RA, Hauptmann M, et al. (2010) The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. Breast Cancer Res Treat 120: 655-661.

33. Drukker CA, Bueno-de-Mesquita JM, Retèl VP, van Harten WH, van Tinteren H, et al. (2013) A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. Int J Cancer 133: 929-936.

34. Mittempergher L, de Ronde JJ, Nieuwland M (2011) Gene expression profiles from formalin fixed paraffin embedded breast cancer tissue are largely comparable to fresh frozen matched tissue. PLoS One 6: e17163.

35. Marrone M, Stewart A, Dotson WD (2015) Clinical utility of gene-expression profiling in women with early breast cancer: an overview of systematic reviews. Genet Med 17: 519-532.