ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 9, Issue 6, 2016

Online - 2455-3891
Print - 0974-2441

Research Article

TRANSCRIPTOME ANALYSIS OF mRNA EXPRESSION PROFILING TO PREDICT POTENTIAL BIOMARKERS INDUCED IN BREAST CANCER-BRCA1

RESHMA JACOB J[1]*, SEEMA J PATEL[1], PRASHANTHA CN[2]

[1]Department of Biotechnology, GM Institute of Technology, Davanagere, Karnataka, India. [2]Founder Chairman and Research, Scientific Bio-Minds, Bangalore, Karnataka, India. Email: reshu0220@gmail.com

Received: 29 July 2016, Revised and Accepted: 10 August 2016

ABSTRACT

Objective: Breast cancer (BRCA1) is the most common cancer type to women with 5,00,000 deaths in every year worldwide. The purpose of the present study is to predict the somatic variants and copy number variants present in BRCA1 genes using 6 paired samples of the glandular tissues.

Methods: In this context, DNA sequencing technology is advanced method used to identify biomarkers that helps patients to have treatment.

Results: From whole genome sequence, several gene mutations are predicted using Annovar and snpEff tools from genome analysis tool kit and classified only non-synonymous single nucleotide variants from all datasets and predicted the 28 novel genes that significantly associated with BRCA1.

Conclusion: Results are classified with both breast cells and BRCA1 cells, identified 13 genes which are most commonly expressed and is further used for molecular diagnostics.

Keywords: Breast cancer, BRCA1, Non-synonymous single nucleotide variants.

 DOI: http://dx.doi.org/10.22159/ajpcr.2016.v9i6.14378

INTRODUCTION

Breast cancer is the most common type of cancer among all other cancer types in women [1]. It is been observed that many etiological factors are implicated in pathogenesis of breast cancer (BRCA1) [2]. Estrogen also acts as a risk factor for breast tumour. Fat rich diet and well-cooked meat increase the incidence of BRCA1 [3]. Obesity also plays an important role in developing BRCA1 [4]. There are several germline gene mutations such as BRCA1 gene also involved in BRCA1 [5,6]. Other genes such as p53, PTEN, CHEK2, and ATM account for a small proportion of hereditary BRCA1 [7]. RNA-Seq, which has a number of edges over conventional expressed sequence tag sequencing. Thus, RNA-Seq experiments not only encapsulate the transcriptome and also used for measuring expression by replacing conventional microarray experiments. Transcriptome sequencing (RNA-Seq) possesses many advantages compared to traditional microarray technology.

The study reveals the feasibility and readiness of comprehensive genetic testing of the BRCA1 gene in an next-generation sequencing (NGS) platform for BRCA1 diagnostics. NGS stage is cost effective and time powerful and meets the affectability and specificity needs requested for hereditary diagnostics, giving NGS and bioinformatics approaches arranged to substitute awesome nuclear instruments in endless genetic diagnostics setups. Therefore, we have included this in our BRCA1 data analysis for disease workflow.

METHODS

Bioinformatics methods are used to analyze the RNA-Seq data from different steps such as pre-processing, quality analysis, gene annotation, differential gene expression analysis, and variant analysis. Datasets are downloaded from ENA and sequence read archive databases. These datasets are subjected to quality analysis using FastQC tool. Tophat tool is used for gene annotation. Results of Tophat are used to predict differential gene expression by Cufflinks. Genome analysis tool kit (GATK) is used for variant analysis.

RESULTS AND DISCUSION

The experimental sample of BRCA1 patients with invariant glandular tissue and primary infected lymph tumor MRNA sequence is used to sequence using the IlluminaMiSeq platform. The overall experiment has paired reads of the 6 samples of the glandular tissues is used to predict the somatic variants and copy number variants present in BRCA1 gene. Further, we need to identify the associated gene functions which are expressed in the BRCA1 tissues. The overall raw RNA-Seq datasets is listed in Table 1.

FastQC tool I used predict the quality using basic statistical quality analysis methods such as mean and median calculation of the box plot prediction. During quality filtering of Illumina data generates 65% of passed clusters within accepted ranges produce more clusters within median range of 14,556,088 to 16,882,897 within the median values 15,794,284. There is no significant difference in the each replicate with statistical *t*-test of 0.5. The overall sequence data is used further used for sequence annotation.

The resultant FastQC trimmed datasets are used to map with reference genome sequence (hg18) to ordinate all nucleotides. The overall compact of TopHat of multiple reads displays input information has 22799 sequence read inputs is mapped with reference genome of 17655 (77.4% of input). Further, cufflinks are used to predict the transcriptional regulated genes of exon and functional transcripts shows 28,856 exons and 36,185 transcripts are predicted from whole genome. To predict total differential expressions, transcript variants upregulation and down regulated genes shows 48 gens were breast tumor BRCA1 associated genes (Table 3a, b). Based on the differential expression of observed Cuffdiff results have three ranges of differential expressions CDS overloading differential expressions, promoters differential expressions and transcriptional differential expressions. We have predicted 1220 upregulated genes and 6858 down regulated genes were extracted from all three types of studies.

Using variant analysis of selected upregulated and downregulated genes to predict clinical variants of BRCA1 is predicted using Annovar

Table 1: RNA - Seq read of glandular cells and primary lymph nodes in breast cancer patients

Run	Spots	Bases	Spots	Averge lengths	Size MB	Library name	Library layout	Sex
ERR687876	22,799	1,14,29,385	22,799	501	5	915 H	Paired	Female
ERR687877	29,090	1,45,86,611	29,090	501	7	916 H	Paired	Female
ERR687878	24,454	1,22,59,739	24,454	501	5	903 H	Paired	Female
ERR687879	34,341	1,72,17,665	34,341	501	8	608 H	Paired	Female
ERR687880	26,392	13,231,720	26,392	501	6	924 H	Paired	Female
ERR687881	35,737	1,79,17,269	35,737	501	8	642 H	Paired	Female

Table 2: Quality control of RNA-Seq Data using FASTQC server

Parameters	1	2	3	4	5	6	7	8	9	10
PBSQ	38	38	38	38	38	38	38	38	38	38
PTSQ%	100	100	100	100	100	100	100	100	100	100
PSQS	15,755	15,755	14,379	15,755	15,755	14,379	15,755	15,755	14,379	15,755
OS	0	0	0	0	0	0	0	0	0	0
Adapter	1-239	1-229	1-229	1-245	1-229	1-239	1-229	1-239	1-239	1-229
K-mer	7.86	1.63	9.84	1.483	0	7.49	0	6.49	5.38	0

SNVs: Single nucleotide variants

Table 3a: Top upregulated genes in breast cancer glandular tissues

Gene names	Term	p	Count	%	Fold enrichment	FDR
KLHL17	Positive regulation of apoptosis	3.74E-05	43	4.5842	1.9634	0.0667
GPR153	Positive regulation of programmed cell death	4.37E-05	43	4.5842	1.9498	0.0779
PEX14	Positive regulation of cell death	4.86E-05	43	4.5842	1.9408	0.0866
FAM131C	Induction of apoptosis	2.02E-04	33	3.5181	2.0247	0.3594
Q13209	Induction of programmed cell death	2.13E-04	33	3.5181	2.0184	0.3803
MST1P9	Regulation of apoptosis	0.00,27,447	60	6.3965	1.4652	4.7871
CROCC	Regulation of programmed cell death	0.00,34,07,136	60	6.3965	1.4508	5.9096
HSPG2	Regulation of cell death	0.00,37,47,257	60	6.3965	1.4454	6.4811
MYOM3	Induction of apoptosis by extracellular signals	0.02,71,00,235	12	1.2793	2.1036	38.7597
RCAN3	Programmed cell death	0.15,33,98,078	38	4.0511	1.2211	94.8807
RHD	Apoptosis	0.17,53,08,035	37	3.9445	1.2067	96.7940

Table 3b: Down regulated genes expressed in breast cancer

Gene name	Term	p	Count	%	Fold enrichment	FDR
NBPF10	Actin-binding	6.54E-04	25	2.6652	2.1464	0.9439
ARHGAP29	Actin binding	0.0,01,72,645	31	3.3049	1.8263	2.6608
GCLM	Cytoskeleton	0.00,43,40,229	46	4.9040	1.5338	6.1119

SNVs: Single nucleotide variants

and snpEff tools from GATK predicts several gene mutations from whole genome sequence and classified only non-synonymous single nucleotide variants from all datasets and extracted to predict the novel genes that significantly associated with BRCA1.

From extracted data identified, 28 novel genes is used to predict functional annotation and enrichment shows all genes have apoptotic character and also associated with certain cancer pathways. Using biomarkers, identification of 28 genes is further screened based on the functional enrichment analysis and classify with both breast cells and BRCA1 cells showed only 13 genes.

Kluska *et al.* had used the TaqMan SNP genotyping test, the rate of 5 favored changes, one normal pleomorphism (Q356R), and 4 varieties (Q563X, N3124I, c.9118-2G, and c.7249delCA >A) were examined in additional gathering of 445 tumor patients analyzed. Despite the fact that Q356R was recognized as morbific in ClinVar, its pervasiveness was indistinguishable among patients (15.1%) and salubrious ascendances (16.9%). By complexity, Q563X was found in 5 (1.1%) out of 445 tumor patients, separately; however, just in one salubrious person. The 3 different varieties (c.7249delCA, c.9118-2G, and N3124I >A) were found by TaqMan genotyping just in DNA tests in which they were recognized by NGS [8]. In our study, we have identified 13 genes which is most commonly expressed and is further used for molecular diagnostics.

CONCLUSION

Our study highlights the differential expression of genes which expressed in different tissues of breast cells is predicted using RNA-Seq analysis. We have characterized the transcriptional factors and differential expression of BRCA1 gene that significantly express in different cell types. This study reveals the feasibility and readiness of

Table 4: BRCA1 genetic mutations is extracted based on non synonymous SNV

Variant type	Functional_annotation	Gene	dbSNP	Phred scaled quality	Quality
Non synonymous	NM_007297:exon22 NM_007298:exon22 NM_007294:exon23 NM_007300:exon24	BRCA1	rs201196020	55.05	3.72
-	-	BRCA1	rs8176297	60	22.98
-	-	BRCA1	rs3092994	60	16.82
-	-	BRCA1	rs8176235	60	12.27
-	-	BRCA1	rs8176234	60	11.24
-	-	BRCA1	rs8176233	60	8.12
Non synonymous SNV	NM_007297:exon14 NM_007298:exon14 NM_007294:exon15 NM_007299:exon15 NM_007300:exon16	BRCA1	rs1799966	59.87	16.04
-	-	BRCA1	rs8176212	58.89	13.35
-	-	BRCA1	rs8176194	58.97	12.15
-	-	BRCA1	rs8176193	56.3	14.5
-	-	BRCA1	rs4793197	58.17	24.1
Synonymous SNV	NM_007297:exon11 NM_007298:exon11 NM_007294:exon12 NM_007299:exon12 NM_007300:exon12	BRCA1	rs1060915	59.71	15.57
Non synonymous	NM_007297:exon9 NM_007294:exon10 NM_007300:exon10	BRCA1	rs16942	60.37	14.28
Non synonymous	NM_007297:exon9 NM_007294:exon10 NM_007300:exon10	BRCA1	rs16941	59.88	16.19
Non synonymous	NM_007297:exon9 NM_007294:exon10 NM_007300:exon10	BRCA1	rs799917	59.96	16.9
synonymous	NM_007297:exon9 NM_007294:exon10 NM_007300:exon10	BRCA1	rs16940	59.88	15.78
synonymous	NM_007297:exon9 NM_007294:exon10 NM_007300:exon10	BRCA1	rs1799949	60	15.83
-	-	BRCA1	rs8176145	60	9.42
-	-	BRCA1	rs8176140	58.43	8.32
-	-	BRCA1	rs799912	60	8.11
-	-	BRCA1	rs3765640	60	20.86

SNVs: Single nucleotide variants

Table 5: Novel gene prediction from breast cancer data to predict potential biomarkers

Data sets	Number of mutations	Non-synonymous	Novel genes	Gene names
1	23,737	232	20	HRNR, KIF9, GPAA1, TONSL, AC012621.2, SULT1A1, FAM92B, TBC1D3C, ZNF208, SEZ6L
2	22,838	541	25	AGBL2, GPD1, SIGLEC6, SEZ6L
3	24,483	229	13	CSF1, HRNR, CCDC19, PDE6B, TSG101, AGBL2, RPLP0, VPS13C, FAM92B, MYH4, TBC1D3C, ZNF208, SEZ6L
4	23,022	186	12	HRNR, SCRIB, DNAH9, TBC1D3C, ZNF208. SEZ6L. HSCB
5	23,586	215	13	ADPRHL2, EIF2B3, HRNR, AAMP, BRPF1 GOLGA8DP, SEZ6L
6	23,352	574	11	PRAMEF2, HRNR, AAMP, CCDC108, HYAL3, ALAS1, RSPH10B2, FAM92B, DNAH9, ZNF208, SEZ6L

genetic testing of BRCA1 associated genes that demand for predicting for clinical diagnostics to detect early stage of infection. Finally, we have classified the genes that are present in both normal breast cells and somatic cell mutation which cause breast tumor.

REFERENCES

1. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, *et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: A combined analysis of 22 studies. Am J Hum Genet 2003;72(5):1117-30.
2. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15(1):17-35.
3. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: Reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94(8):606-16.
4. Rosen EM, Fan S, Pestell RG, Goldberg ID. BRCA1 gene in breast cancer. J Cell Physiol 2003;196(1):19-41.

5. James CR, Quinn JE, Mullan PB, Johnston PG, Harkin DP. BRCA1, a potential predictive biomarker in the treatment of breast cancer. Oncologist 2007;12(2):142-50.
6. Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell 2002;108(2):171-82.
7. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266(5182):66-71.
8. Kluska A, Balabas A, Paziewska A, Kulecka M, Nowakowska D, Mikula M, *et al.* New recurrent BRCA1/2 mutations in Polish patients with familial breast/ovarian cancer detected by next generation sequencing. BMC Med Genomics 2015;8:19.