

RESEARCH ARTICLE

Open Access



Down regulation of KLK7 expression in breast tissues and identification of a novel spliced KLK7 mRNA

Samina Ejaz¹, Faiz-ul-Hassan Nasim^{2*}, Muhammad Ashraf² and Gulzar Ahmad³

Abstract

Background: Alternative splicing commonly occurs in cancer cells and many cancer specific splice variants have been reported as potential candidate biomarkers of the disease. We have studied human tissue Kallikrein 7 (KLK7) mRNA expression profile in breast cancer patients of our region. KLK7 is member of a multi-gene family consisting of 15 members (KLK1-KLK15).

Methods: We optimized touch down nested PCR method for the amplification of KLK7 isoforms/variants. Various bioinformatics tools were used for sequence analysis, identification of splicing pattern and prediction of encoded proteins.

Results: We observed an unusual splicing event consisting of exon 3 (E3) truncation at 3' end (by 124 nucleotides), exon 4 (E4) exclusion and exon 5 (E5) truncation at 5' end (by 33 nucleotide) in 2 normal breast tissues, one obtained from invasive ductal carcinoma grade II patient and other collected from mammary dysplasia patient. Moreover, 3 other KLK7 mRNAs (KF963190, KF963191, and KF963193) expressed in breast cancer were noticed to exhibit single nucleotide polymorphism (SNPs). Bioinformatic analysis revealed that the alternatively spliced mRNA (KF963192) will potentially encode a truncated and non-functional protein. Similarly although encoded proteins have considerable homology with normal hK7 protein, SNPs seem to cause great variations in pls, structures and molecular weights of encoded proteins.

Conclusions: There is need to further explore the impact of the unique splicing event, SNPs and characterize these population specific mutations and their possible role in the pathogenesis of breast cancer.

Keywords: Alternative splicing, Biomarker, Breast cancer, KLK7, Splice variants

Background

Alternative splicing is a way to produce functionally diverse proteins from a single gene [1, 2]. It occurs in a variety of biological processes including developmental, physiological and pathogenic pathways [3]. According to one estimate at least 15% of the genetic diseases are the outcome of defective pre-mRNA splicing and alternative splicing is a common phenomenon occurring in cancer cells. Various cancer-specific splice variants have been proposed as potential biomarkers of cancer [4–6].

Human tissue Kallikreins (KLKs) is a multigene family consisting of 15 members which encode serine proteases. All KLKs are known to produce alternative transcripts through the use of alternative, transcriptional start site, polyadenylation and splicing. Browsing of the Genbank database revealed that KLKs have 82 reported transcripts and 49 (~ 27% of 82) are the outcome of alternative splicing. Some of the splice variants are tissue, developmental stage, stimulus or disease specific. Few KLK splice variants help in cancer diagnosis and prognosis [7–9] and can be used as potential biomarkers of cancer [10, 11]. Future work in this area will be particularly helpful to find more useful diagnostic and prognostic biomarkers [12].

* Correspondence: faiz.nasim@iub.edu.pk

²Department of Chemistry, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

Full list of author information is available at the end of the article



KLK7 is a member of KLKs family and is known to encode human stratum corneum chymotryptic enzyme (hSCCE of PRSS6) and expressed in a wide range of tissues [13–15]. According to GenBank, a total of 11 distinct transcripts have been reported for KLK7 (Table 1 and Fig. 1) and only three can be categorized as alternatively spliced transcripts/isoforms. KLK7 mRNA expression level has been quantified by RT-PCR in breast cancer tissues and reported to be a better prognostic marker for the unfavorable prognosis of breast carcinoma. KLK7 mRNA was significantly lowered in either stage I or stage II breast cancer patients and higher level of KLK7 mRNA was found to be associated with better prognosis [16, 17].

Breast cancer is the second leading cause of death worldwide [18]. In Asia Pakistan has the highest prevalence of breast cancer. On the average about 40,000 women die per year due to this disease [19, 20] and one out of every nine women in Pakistan is at the risk of developing breast cancer [21]. Contrary to western females, Pakistani females suffer from the breast cancer at an early age having a higher risk of metastatic cancers and are presented with large lesions [22].

Considering the importance of the issue and association of KLK7 mRNAs with breast cancer we hypothesized that based on genetic differences there can be certain new splice variants expressed in breast cancer (BC) patients of our region. The project was therefore initiated to evaluate the potential of KLK7 isoforms/variants as biomarkers of breast cancer in this region. This study is the first to report KLK7 isoforms/variants in patients of southern Punjab, Pakistan.

Methods

The study was approved by appropriate institutional technical committees including the departmental Ethics and Research Committee and by the Advanced Studies and Research Board (ASRB) of the Islamia University of Bahawalpur, Pakistan. Written Consent to participate in the study was taken on a prescribed form from all patients included in the study. The experimental work involving humans was in complete compliance with the Helsinki Declaration.

Mean age of patients was 47.32 ± 12.16 years (range 24–75 years) and majority of patients were suffering from infiltrating ductal carcinoma (53.20%). Total 118 breast tissue samples, 63 diseased and 55 normal samples were subjected to analysis. Normal tissues were carefully removed by a surgeon from regions away from area of cancerization. However, in case of 8 patients due to the involvement of whole breast it was not possible to obtain normal tissue. Following the instructions of manufacturers tissue samples submerged in RNAlater^R were transported from hospital to laboratory, incubated overnight at 2–8 °C followed by the removal of tissues from RNAlater^R and storage at -70 °C.

Extraction and quantification of RNA

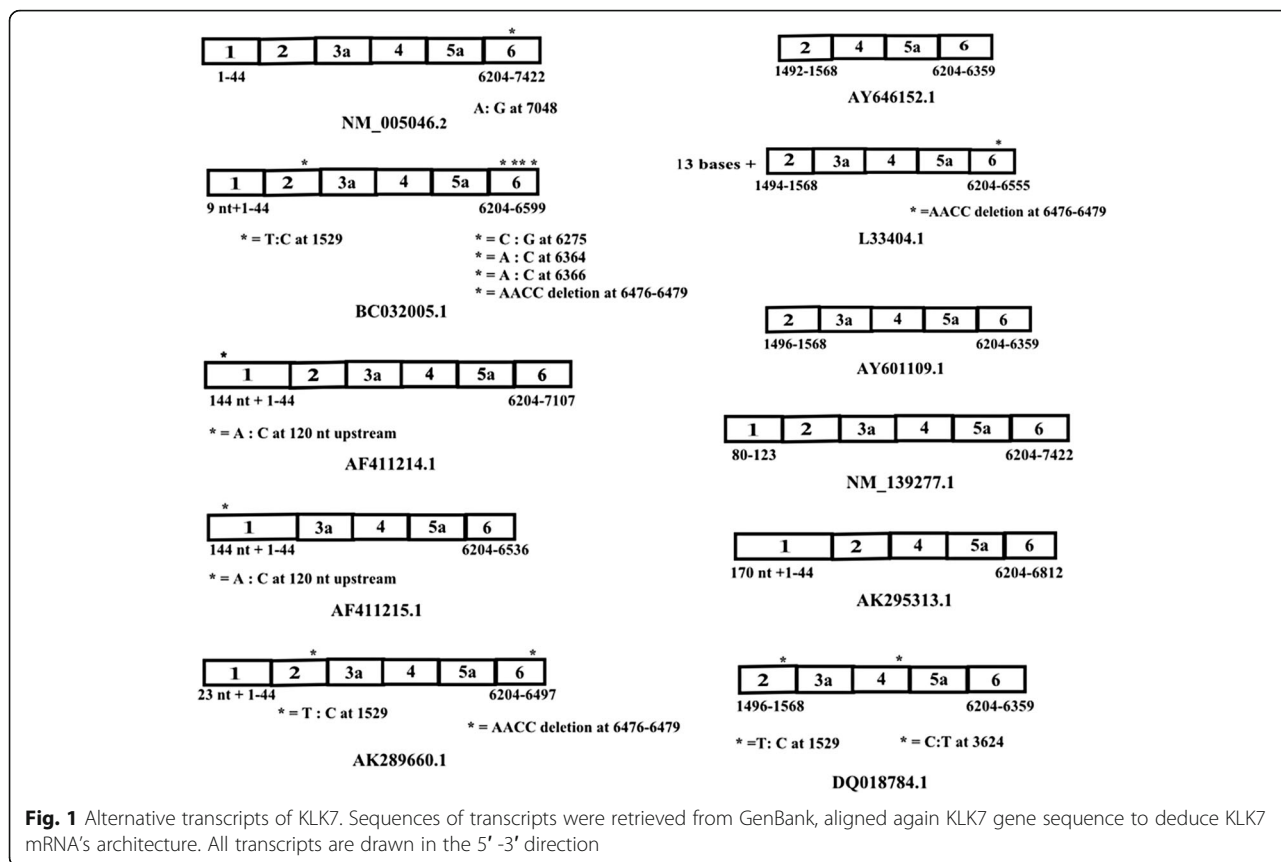
Tissue samples were processed for the extraction of total RNA using Norgen Biotek corporation 'total RNA purification kit' (Catalogue No. 17200). RNA was quantified by measuring absorbance (A) at 260 nm and purity of RNA was checked by measuring the ratio of A260/280. RNA samples having A260/280 ratio between 1.8–2.1 were further subjected to PCR amplifications.

cDNA synthesis

cDNA was synthesized by reverse transcription using M-MLV (Moloney Murine Leukemia Virus) reverse transcriptase (Invitrogen, catalogue no. 28025–013). Reverse transcription was carried out in a reaction mixture of 20 µl. Initially RNA solution containing 3 µg RNA, 1.5 µl of 20 µM poly dT primer (1.5 µM), 1 µl of 10 mM dNTPs mix (0.5 mM) were mixed in 0.2 ml PCR tube placed on ice. Total volume was adjusted to 12.5 µl with nucleases free H₂O, if required. Mixture was heated at 65 °C for 5 min and quickly chilled on ice. After cooling the contents of tube were collected by brief centrifugation. It was followed by addition of 6.5 µl RT mix containing 4 µl of 5× first strand buffer (supplied by Invitrogen with M-MLV reverse transcriptase), 2 µl 0.1 MDTT (10 mM) and 0.5 µl nucleases free H₂O. Contents of tubes were gently mixed by pipetting and incubated at 37 °C for 2 min. Finally 1 µl (200 units) M-MLV reverse transcriptase (M-MLV RT) was added to the reaction mixture, contents were gently mixed by

Table 1 Already known KLK7 alternative transcripts

Sr. no.	Constitutively spliced mRNAs	Constitutively spliced mRNAs having SNP	Alternatively spliced mRNAs	Alternatively spliced mRNAs having SNP
1	AY601109.1	NM_005046.2	AY646152.1	AF411215.1
2	NM_139277.1	BC032005.1	AK295313.1	-----
3	-----	AF411214.1	-----	-----
4	-----	AK289660.1	-----	-----
5	-----	L33404.1	-----	-----
6	-----	DQ018784.1	-----	-----



pipetting up and down and incubated at 37 °C for 50 min. After 50 min of incubation at 37 °C M-MLV RT was inactivated by heating at 70 °C for 15 min and the reaction mixture was stored at either 4 °C (for short term storage) or -70 °C for long term storage.

Designing/selection of primers for PCR amplification

Primers for PCR amplification of KLK7 and β-actin (used as internal control) cDNAs were selected from a list of possible suitable primers predicted by the primer prediction software Primer Premier 6.0. Amplification of cDNA corresponding to full length KLK7 mRNA was main criteria used for primer selection. Since nested PCR protocol was used during the present study two pairs of primers were used for KLK7 mRNAs expression studies. One pair to amplify exon 1 through exon 6 and second pair to amplify exon 2 through exon 6 region of KLK7 cDNAs (Table 2).

PCR method used for KLK7 and β-actin amplification

For PCR amplification of KLK7 we optimized touch down (TD)-nested PCR method. To check the integrity of RNA and normalize expression of KLK7, β-actin gene was used as internal control. Touch down PCR (TD-PCR) is relatively simple and rapid method with enhanced sensitivity, specificity and yield of amplification.

In this technique annealing temperature is initially kept above the melting temperature of primers being used for amplification and then over the course of successive cycles transitions are progressively made to a lower annealing temperature. TD PCR is better employed to amplify transcripts which are difficult to amplify. Moreover, TD-PCR is helpful for rapid optimization of PCR conditions avoiding lengthy experimentation for optimization [23].

In nested PCR 2 sets of primers are used and PCR amplification is done in two successive runs of PCR. First run of PCR is carried out using one set of primers and during second run second set of primer is used which hybridizes or targets regions within the first run product [24, 25]. To amplify cDNAs corresponding to KLK7 mRNAs we combined TD-PCR with nested PCR technique to enhance chances of amplification. First round of PCR was performed using gene specific primers F1 and R1 and second round was carried out by employing primers F2 and R2 targeting the region within the amplified product of 1st round. Constitutive splicing of KLK7 pre-mRNAs resulted in the production of 697 bp PCR amplified product.

Touch down-nested PCR for KLK7 cDNAs amplification

In the first round of PCR 5 μL of RT-mixture was taken in a 0.2 ml PCR tube and 20 μL of PCR master mixture

Table 2 Characteristics of primers used to study KLK7 expression

Sr. no.	Gene	Primer	Position of hybridization (Exon: position)	Sequence of primer	T _m (°C)	Length (nt)	Nature of primer
1	KLK7	F1	E1: 11–28	AGTCGGAACCTGGATCAC	52.5	18	sense
2		R1	E6: 6356–6373	AACTCAGTGTGGCGTTAG	53.0	18	antisense
3		F2	E2:1532–1550	CTGCTATCCTTAGCCTTGG	53.5	19	sense
4		R2	E6:6313–6330	TCCACTTGGTGAACCTTGC	53.6	18	antisense
5	β-actin	F1	E4: 2253–2271	TGCGTGACATTAAGGAGAA	51.5	19	sense
6		R1	E4: 2407–2424	AAGGAAGGCTGGAAGAGT	53.9	18	antisense

was added. PCR master mixture contained 2.5 µL of 10X PCR buffer (without MgCl₂), 0.5 µL of 10 mM dNTPs solution, 2 µL of 25 mM MgCl₂, 1 µL of 20 µM each primer solution, 0.25 µL of 5 U/µL Taq DNA polymerase, 12.75 µL Nucleases free H₂O to get concentration of 1X PCR buffer (without MgCl₂), 200 µM dNTPs, 2 mM MgCl₂, 20 pmoles of each primer, 1.25 U of Taq DNA polymerase in the final PCR reaction mixture (total volume 25 µL).

Thermal cycler was programmed to run 5 segments. Segment 1 consisted of heating of PCR reaction at 95 °C for 5 min. Segment 2 comprised of 3 PCR steps (i.e., denaturation at 94 °C for 30 s, annealing at 63 °C for 1 min, and extension at 72 °C for 1 min). Segment 2 was iterated for 10 cycles with a decrement of 1 °C in annealing temperature, starting from 63 °C (~ 10 °C above T_m) and ending up at 54 °C. Segment 2 was followed by segment 3 which also consisted of 20 cycles of 3 PCR steps (i.e., 94 °C for 30 s, 54 °C for 1 min, and 72 °C for 1 min). Segment 4 included final amplification step at 72 °C for 15 min. Finally the reaction was held at 4 °C (segment 5) and proceeded for 2nd round of PCR.

For 2nd round of PCR, master mix was prepared by mixing the same volume of different components as in 1st round of PCR except that 14.25 µL Nucleases free H₂O was added instead of 12.75 µL to get a total volume of 22.5 µL and primers F1R1 were replaced with primers F2R2. To this PCR master mix 2.5 µL of 1st round PCR mix was added and tube was placed in thermal cycler which was programmed for segment 1: 94 °C for 10 min, segment 2: 25 cycles of 3 PCR steps (94 °C for 30 s, 55 °C for 1 min and 72 °C for 1 min) along with final amplification at 72 °C for 15 min (segment 3). Reaction was finally held at 4 °C (segment 4) and 5 µL of 2nd round PCR reaction was loaded on 2% agarose gel to resolve amplified PCR products.

PCR method used for β-actin amplification

To amplify β-actin cDNA 2.5 µL of RT, 22.5 µL of PCR master mixture containing 1X PCR buffer (without MgCl₂), 200 µM dNTPs, 2 mM MgCl₂, 20 pmoles of each primer, 1.25 U of Taq DNA polymerase and Nucleases free H₂O were mixed to make total volume of

25 µL. PCR reaction mixture was held at 95 °C for 5 min followed by 30 cycles of 3 PCR steps (step1: 94 °C for 30 s, step 2: 51 °C for 1 min, step 3: 72 °C for 1 min) and final amplification step (72 °C for 15 min). Reaction was finally held at 4 °C (segment 4) till further processing.

Sequencing of PCR amplified products

PCR amplified products were resolved on 2% gels. Amplified products were cut from the gel and sent for sequencing to MacroGen Incorporation, Seoul, Korea. Sequences obtained were aligned against KLK7 genomic sequences retrieved from GenBank database and mRNA sequences were deduced.

Bioinformatics analysis

Bioinformatics analysis was performed using online databases and analysis tools. To reduce chances of false positive and false negative results two precautionary measurements were made during the study. First, we performed visual inspection followed by computational analysis and second, in most of the cases we used multiple softwares.

Retrieval of genes and transcript sequences

Sequences of the genes and reported transcripts were retrieved from NCBI 'gene' (<http://www.ncbi.nlm.nih.gov/gene>) and nucleotide (<http://www.ncbi.nlm.nih.gov/nucleotide>) databases and analyzed using different bioinformatics application packages (Table 3). This information was used to build corresponding pre-mRNAs by inserting intronic sequences at the junction site between the exonic regions.

Molecular weights, theoretical pI of the predicted proteins were calculated using ExPASy 'Compute pI/MW tool'. For predicting sub-cellular localization of proteins softberry tool 'ProtComp v. 9.0' was employed. Three dimensional structures of proteins were modeled using 'Phyre 2' which also helped to determine homology index of proteins.

KLK7 mRNAs sequences being expressed in the breast tissues used during the present study were submitted to GenBank database (Accession numbers

Table 3 Softwares used for the computational analysis of transcripts and proteins

Purpose	Softwares used	Available at:
Sequence alignment analysis	EMBOSS Pairwise Alignment Algorithms	http://www.ebi.ac.uk/Tools/psa/
Splice site prediction, determining splice site strength and confidence level	SpliceView	http://bioinfo.itb.cnr.it/oriel/splice-view.html
	Neural Network	http://www.fruitfly.org/seq_tools/splice.html
	NetGene 2	http://www.cbs.dtu.dk/services/NetGene2
	ESE-Finder	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi
	Human Splicing Finder (HSF) version 2.4	http://www.umd.be/HSF
Finding splicing enhancers and splicing silencers	ESE-Finder	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi
	Human Splicing Finder (HSF) version 2.4	http://www.umd.be/HSF
mRNA sequence conversion to protein	Translator	http://www.fr33.net/translator.php
Calculating Molecular weight and pI of protein	ExPASy 'Compute pI/MW tool'	http://web.expasy.org/compute_pi/
Predicting sub-cellular localization of proteins	ProtComp v. 9.0	http://linux1.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc
Prediction of 3D structure and determination of homology index	Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

KF963189- KF963193). The data has been made available online (<https://www.ncbi.nlm.nih.gov/gene/5650>).

Sequence alignment, transcript analysis and identification of SNPs

Similarities or differences between the KLK7 mRNAs being expressed in breast cancer tissues and already known constitutively spliced mRNAs were deduced through sequence alignment using EMBOSS Pairwise Alignment Algorithms. Transcripts having variations at the internal sequences were designated as alternatively spliced isoforms. Pairwise needle alignment also helped to identify existing SNPs, if any. Deduced pre-mRNAs were subjected to splice site prediction and analysis tools (Table 3) and distribution of exonic/intronic splicing enhancer (ESE/ISE) and exonic/intronic splicing silencer sequences (ESS/ISS) were determined using ESE (Exonic Splicing Enhancer) and HSF (Human Splicing finder) softwares.

Prediction of protein products

On the basis of earlier known information protein coding sequences of mRNAs were marked and translated into protein sequences using software 'Translator' which can potentially convert mRNA sequences into proteins.

Proteins encoded by known constitutively spliced mRNAs of corresponding genes were taken as reference proteins. All best possible frames of proteins encoded by the mRNAs being expressed in breast cancer tissues were aligned using EMBOSS pairwise alignment algorithms to the reference protein isoforms to know the

extent of homology between proteins encoded by already reported mRNAs and mRNAs being expressed in breast cancer patients included in the present study.

Results

Total 118 breast tissue samples, 63 diseased and 55 normal samples were subjected to analysis. Surgeon carefully removed normal tissues in 55 patients from region away from area of cancerization. However, in case of 8 patients due to the involvement of whole breast it was not possible to obtain normal tissue. Only 6 (3 normal, 2 cancerous and 1 benign) of 118 samples showed PCR amplification when amplified using KLK7 specific primers. Hence the mRNA positivity ratio observed in our patients was even lower than that has been reported earlier [17]. We noticed expression of 3 different types of KLK7 transcripts (~700, ~300 and ~200 bp) in breast cancer patients of our study (Fig. 2). PCR amplified products resolved on agarose gel were named as SER1-SER5. To deduce KLK7 mRNAs sequences of the amplified products were aligned against KLK7 gene's sequence (GenBank Accession no. NC_000019.10). Results of alignment analysis indicated expression of both constitutive and alternatively spliced mRNAs in different breast cancer tissues (Fig. 2 and Table 4). Constitutively spliced mRNAs (SER1-SER3, SER5) were detected in 3 tissues (6 N, 18 N and 18C). However, two breast tissues (16 N and 29 N) exhibited expression of alternatively spliced KLK7 mRNA (SER4) and in only one breast tissue (16C) both the constitutive and the alternative types of KLK7

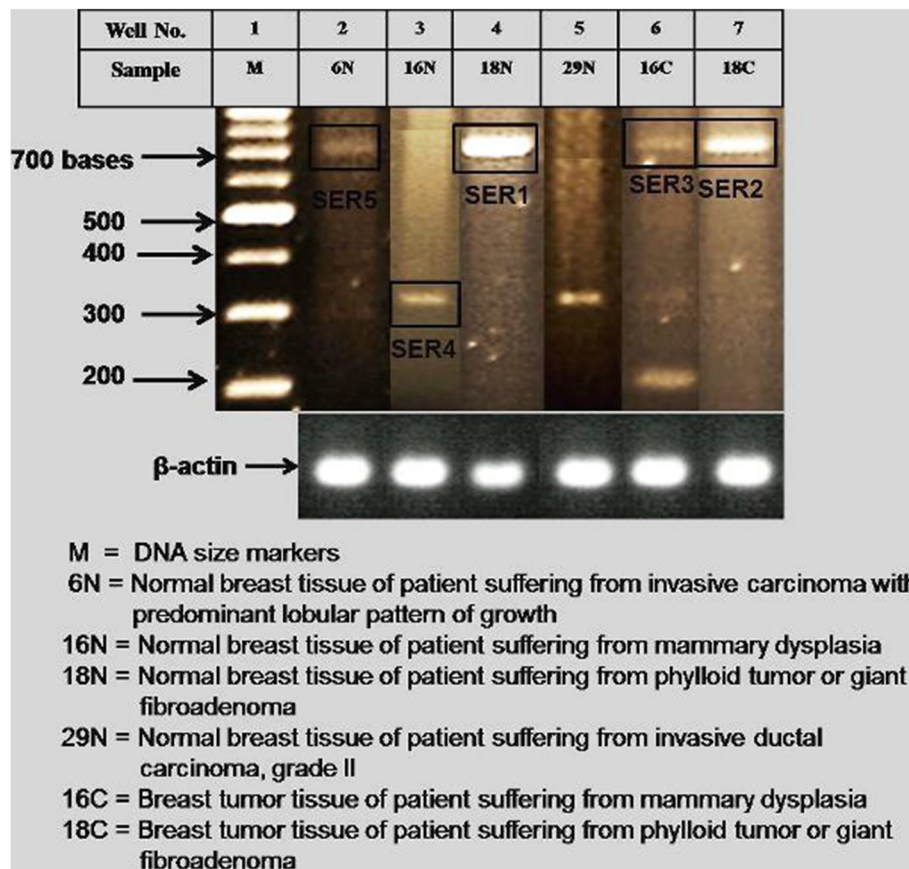


Fig. 2 PCR amplified products resolved on agarose gel. RNA was extracted, reverse transcribed using poly dT primers, cDNA amplified using KLK7 primers through touch down nested PCR method. M = DNA size markers. 6 N = Normal breast tissue of patient suffering from invasive carcinoma with predominant lobular pattern of growth. 16 N = Normal breast tissue of patient suffering from mammary dysplasia. 18 N = Normal breast tissue of patient suffering from phylloid tumor or giant fibroadenoma. 29 N = Normal breast tissue of patient suffering from invasive ductal carcinoma, grade II. 16C = Breast tumor tissue of patient suffering from mammary dysplasia. 18C = Breast tumor tissue of patient suffering from phylloid tumor or giant fibroadenoma

mRNA splicing was observed (Accession No. KF963189-KF963193, Fig. 3 and Table 5).

Analysis of constitutively spliced KLK7 mRNAs

For identification of splicing scheme already known constitutive splicing scheme was used as reference (Fig. 4). Out of 4 constitutively spliced KLK7 mRNA expressed in breast cancer patients of our study only one mRNA (SER1) was exactly similar to the already known constitutively spliced mRNA (NM_005046.2) taken as standard

(Figs. 3 and 4). However, other 3 constitutively spliced mRNAs contained single nucleotide polymorphisms (SNPs).

Predicted encoded proteins

Predicted protein encoded by mRNA1/NM_005046.2 was taken as reference protein isoform. Using already reported information about protein coding region of KLK7 mRNA (last 73 nucleotides of E2, E3, E4 and 156 nucleotides of E5), encoded protein sequences were deduced

Table 4 Nature of KLK7 splice products expressed in breast cancer patients

Sr. no.	Nature of spliced product	Total (n = 118)		Normal (n = 55)		Diseased (n = 63)	
		No.	Breast tissue	No.	Breast tissue	No.	Breast tissue
1	Constitutive	3	6 N, 18 N, 18C	2	6 N, 18 N	1	18C
2	Constitutive & Alternative	1	16C	—	—	1	16C
3	Alternative	2	29 N, 16 N	2	29 N, 16 N	—	—
	Sum		6	4		2	

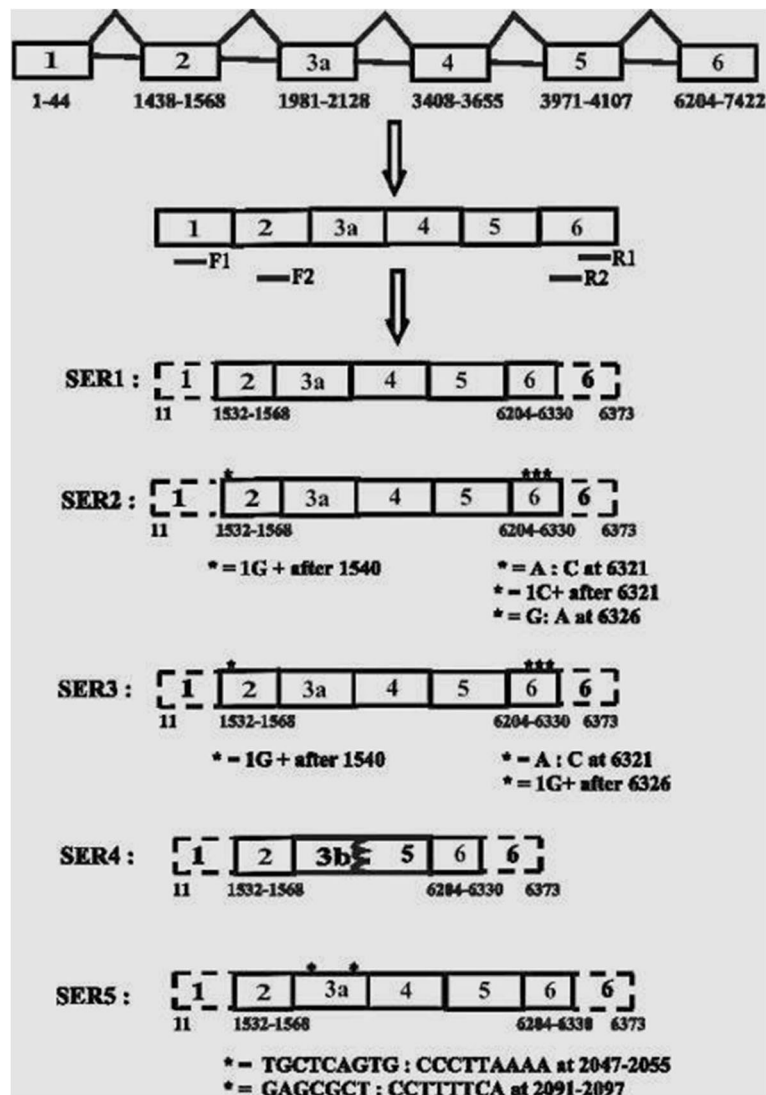


Fig. 3 KLK7 mRNAs expressed in breast tissue samples analysed during present study. Sequences have been submitted to GenBank. All transcripts are drawn in the 5' -3' direction

for 4 constitutively spliced KLK7 mRNAs (Additional file 1: Table S1). Predicted protein encoded by SER1 exhibited perfect alignment with the reference protein (hK7). In case of SER2, 3 and 5 three ORFs were predicted by the software (Translator). Only first 16 amino acids of SER2, and 3 (frame 1) were found to

be perfectly aligned with the first 16 amino acids of the reference protein. Frame 2 was noticed to have considerable alignment (16-24 amino acids) with the reference protein but no perfectly aligned region was observed in case of frame 3 except last 11 amino acids (242–253) in case of SER2 and last 10 amino

Table 5 An overview of KLK7 alternative transcripts expressed in breast cancer patients

Sr. no.	Constitutively spliced mRNAs				Alternatively spliced mRNAs			
	Code	GenBank Accession No.	Size (bases)	Expressed in tissue	Code	GenBank Accession No.	Size (bases)	Expressed in tissue
1	SER1	KF963189	697	18 N	SER4	KF963192	292	16 N, 29 N
2	SER2 ^a	KF963190	697	18C	----	----	----	----
3	SER3 ^a	KF963191	697	16C	----	----	----	----
4	SER5 ^a	KF963193	698	6 N	----	----	----	----

^aIndicates SNPs

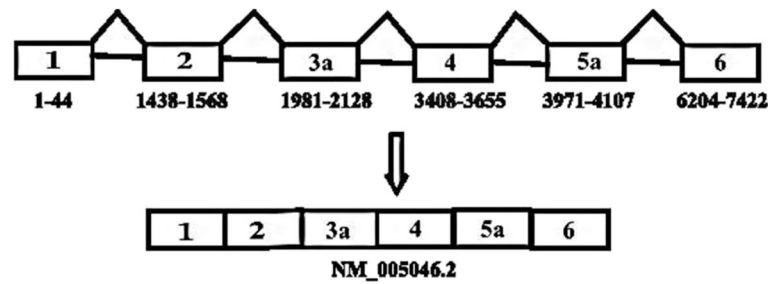


Fig. 4 Constitutive splicing scheme of KLK7 mRNAs deduced by aligning largest size KLK7 mRNA (NM_005046.2) available in GenBank database with KLK7 gene. Sequences were retrieved from GenBank. The transcript is drawn in the 5' -3' direction

acids of SER3 (244–253). Similarly frame 2 of SER5 exhibited significant alignment (65–253) with the reference protein. Partial alignment was observed in SER5 frame 2 but no alignment was seen in case of frame 3 (data not shown). Moreover, multiple premature termination codons were present in frame 1 and 3 of SER 2, 3 and 5 hence short polypeptides would be the translational products.

Three dimensional structure of the protein encoded by SER1 was similar to that of the reference protein (the one encoded by NM_005046.2). However, proteins encoded by SER2 and SER3 were noticed to have three dimensional structures resembling to mannan-binding lectin serine protease1 (Fig. 5). Although both have been categorized as hydrolase, the amino acid sequence and structure was entirely different from hK7. Similarly SER5 encoded protein structure has been found to be closely related with prophenoloxidase activating factor. Theoretical pI (isoelectric point) of proteins ranged from 8.17 to 9.49 (Table 7) and all proteins have been predicted to be secretory ones although trace amounts can be present in other organelles (Additional file 1: Table S2).

Three alternative transcripts of KLK7 expressed in normal (SER5) and cancerous breast tissues (SER2–3) were noticed to exhibit SNPs in protein coding regions (Table 6). Although sequences of the predicted proteins were found to have considerable homology with the reference protein (hK7), the SNPs appear to generate great variations in pI, structure and molecular weights of the encoded proteins (Table 7). Structures of SER2 and SER3 encoded proteins diverge from hK7 and resemble with mannan-lectin serine protease 1 (Fig. 5). Structure of SER5 encoded protein resembled more with plasminogen than hK7.

Analysis of alternatively spliced KLK7 mRNAs

A new splice variant (SER4; KF963192) is being reported that has been found to be expressed in breast tissues. Repetitions of the cDNA synthesis / PCR amplifications consistently resulted in the production of this variant. As for as the mechanism is concerned this transcript has

originated by the combination of three splicing events including 3' truncation of E3, 5' truncation of E5 and exclusion of E4. However, earlier known splicing events have been identified as exclusion of either E2 or E3 (Table 8, Fig. 6). To understand and justify the nature of splicing event, it was essential to know the nature/strength of splice sites and other contributing factors like distribution frequency of splicing enhancer and silencer motifs.

The 5' and 3' splice sites of E3a, E3b, E4, E5a, and E5b were subjected to analysis using various bioinformatics tools. Higher score of splice site is the indicator of better match with the consensus/known sequence of splice sites. Data has been summarized in Additional file 1: Table S3 and S4.

The 3'ss (3' splice site) of E5a (score = 8.145) was observed to be stronger than the 3'ss of E4 (score: 7.388). The presence of upstream 3'ss of E3 (E3b: 1981–2004) and downstream 5'ss of E5 (E5b: 4004–4107) was not supported by any software used for analysis. Upstream 3'ss of E3 is the one that has been reported earlier and downstream 3'ss of E3 used in the splice variant of our study is the unique and has not been reported by any other study. Upstream 5'ss of E3 was not predicted using default threshold values for splice site strength and it was only predicted when threshold value was lowered to 1 and strength of the splice site (E3b: score 2.384) was too low as compared to the splice site used during constitutive splicing (E3a: score 7.986). Similarly the presence of downstream 3'ss of E5 (E5b: 4004–4107) was not predicted by the analysis tools however, another 3'ss (4008–4107) is predicted to be present 5 nucleotides downstream to this splice site. Thus the splicing event of SER4 cannot be justified on the basis of this information.

To explore this aspect alternatively spliced exonic sequences (E3a, E4 and E5a) were analyzed to determine the presence of ESEs/ISEs (SF2/ASF (IgM-BRCA1), SF2/ASF, SC35, SRp40, SRp55, Tra2- β and 9G8) and ESSs/ISSs (Fas, hnRNPA1, Sironi motif 1, 2, and 3). Results indicated highest distribution frequency of all studied ESEs except SRp40 in E3a (Figs. 7). E4 was noticed to have the highest distribution frequency of SRp40 among the

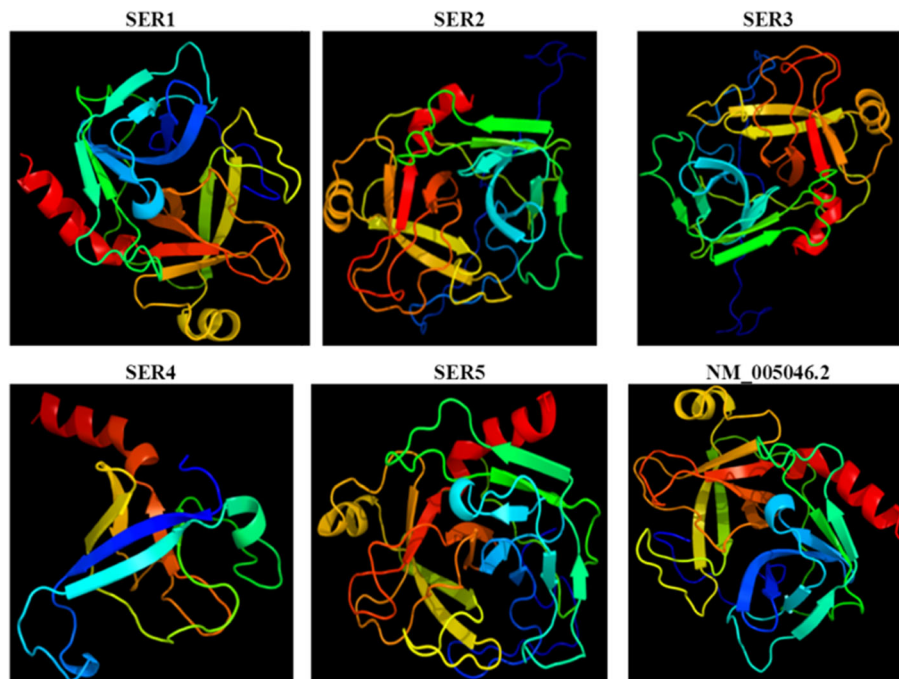


Fig. 5 Predicted/possible 3D structures of KLK7 proteins (obtained through Phyre2 by submitting KLK7 mRNAs sequences). **SER1 PDB Title:** crystal structure of human kallikrein 7 in complex with suc-2 ala-ala-pro-phe-chloromethylketone; **SER2 & SER3 PDB Title:** mannan-binbind binding lectin serine protease 1, structure of a protease **NM_005046.2 PDB Title:** crystal structure of human kallikrein 7 in complex with suc-2 ala-ala-pro-phe-chloromethylketone; **SER4 PDB Title** plasminogen; the x-ray crystal structure of full-length type ii human plasminogen; **SER5 PDBTitle:** crystal structure of prophenoloxidase activating factor-ii from the2 beetle holotrichia diomphalia

three constitutively spliced exons of KLK7. For all other ESEs E4 occupies central position among analyzed exons with the exception of Tra2- β and 9G8 motifs which were noticed to have the lowest distribution frequency in E4. Similarly E3a had the highest distribution frequency of all studied ESSs (Fig. 8), whereas E4 contained the lowest distribution frequency of all studied ESSs except hnRNPA1 binding motif that had the lowest occurrence in E5a.

Protein sequence (Additional file 1: Table S1) deduced from the alternatively spliced mRNA of KLK7 (SER4; KF963192) was subjected to needle alignment against

hk7 protein encoded by NM_005046.2 taken as reference/standard. Results of alignment studies revealed that the protein encoded by SER4 had two perfect regions of alignments including 1–31 and 32–118 amino acids aligned with 1–31 and 67–253 amino acids of reference protein (data not shown).

To have an account of homology between proteins encoded by alternatively spliced isoform expressed in breast cancer tissues during the present study (SER4) and proteins encoded by already reported splice variants (AF411215.1, AY646152.1 and AK295313.1) protein alignment studies were conducted. Results have indicated that

Table 6 SNPs observed in constitutively spliced KLK7 mRNAs expressed in breast cancer tissues

Sr. no.	mRNA	Polymorphism	Position	Effect on sequence of encoded protein	Effect on 3D structure
1	SER2	1C+	after 1540	16–241 amino acids align with the reference protein	Hydrolase, mannan-lectin serine protease1
		A: C	6321		
		1C+	after 6321		
2	SER3	1C+	after 1540	16–241 amino acids align with the reference protein	Hydrolase, mannan-lectin serine protease1
		A: C	6321		
		1G+	after 6326		
3	SER5	TGCTCAGTG: CCCTTAAAA	2047–2055	65–253 amino acids align with the reference protein	Hydrolase, plasminogen
		GAGCGCT: CCTTTTCA	2091–2097		

Table 7 Characteristics of predicted proteins encoded by KLK7 mRNAs

Sr.No.	mRNA encoding protein	Total aminoacids (n)	Molecular weight (Da)	pI
1	SER1	253	27,524.73	8.82
2	SER2	246	26,755.55	8.17
3	SER3	246	26,723.54	8.34
4	SER4	118	12,822.92	7.55
5	SER5	237	26,240.32	9.49
6	NM_005046.2	253	27,524.73	8.82

pI = Theoretical pI

the region consisting of 32–118 amino acids encoded by SER4 was directly aligned to the region 142–228 (AF411215.1) and 117–203 of AY646152.1 and AK295313.1.

According to the protein structure prediction software (Phyre2) the predicted three dimensional structure of SER4/KF963192 encoded protein resembles with plasminogen rather than hK7 (Fig. 5). Similarly theoretical pI was calculated to be 7.55 rather than theoretical pI of native hK7 isoform (8.82). Native/reference protein isoform of hK7 contained 253 amino acids but SER4 encoded protein consisted of 118 amino acids and accordingly its molecular weight varied (Fig. 5, Table 7). However, alternative splicing event did not seem to have any effect on sub-cellular localization of the protein. Like native/reference hK7 isoform SER4 encoded protein has also been predicted to be predominantly extracellular (Additional file 1: Table S2) by protein location prediction software (ProtComp).

Discussion

During the present study KLK7 mRNA expression could be documented in only 6 out of 118 samples hence mRNA positivity ratio was lower than that has been reported earlier [17]. KLK7 mRNA expression was only detected in 6 samples (4 normal, 1 cancerous, & 1

Table 8 An overview of KLK7 alternative transcripts expressed in breast cancer patients of study

Sr. no.	Nature of splice variant	Spliced variants	Splicing event
1	Already reported	AF411215.1	Exclusion of complete E ₂
2		AY646152.1	Exclusion of complete E ₃
3		AK295313.1	
4	Being reported	SER4	Truncation of E ₃ at 3'end by 124 nt Exclusion of complete E ₄ Truncation of E ₅ at 5'end by 33 nt

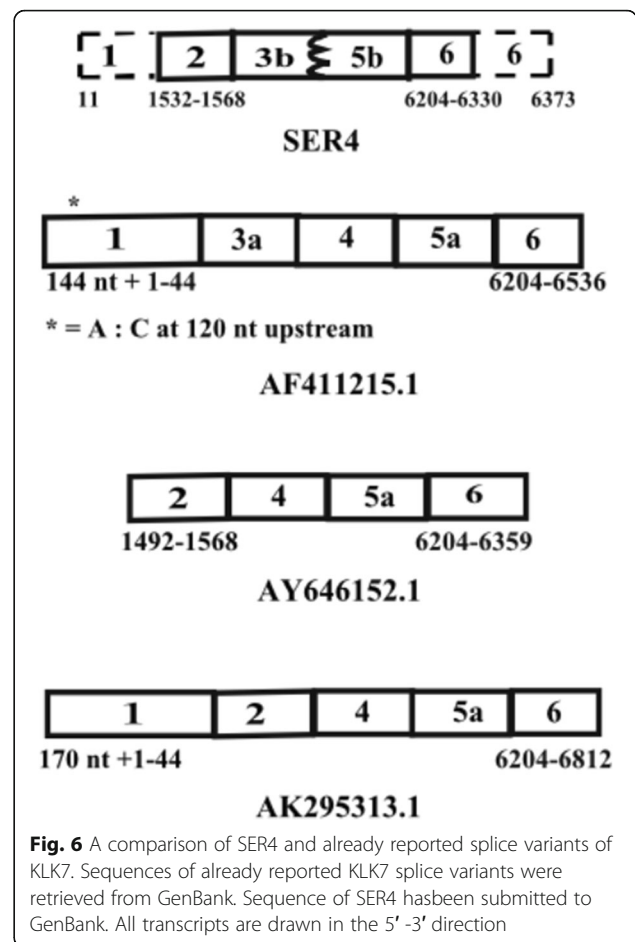


Fig. 6 A comparison of SER4 and already reported splice variants of KLK7. Sequences of already reported KLK7 splice variants were retrieved from GenBank. Sequence of SER4 has been submitted to GenBank. All transcripts are drawn in the 5' -3' direction

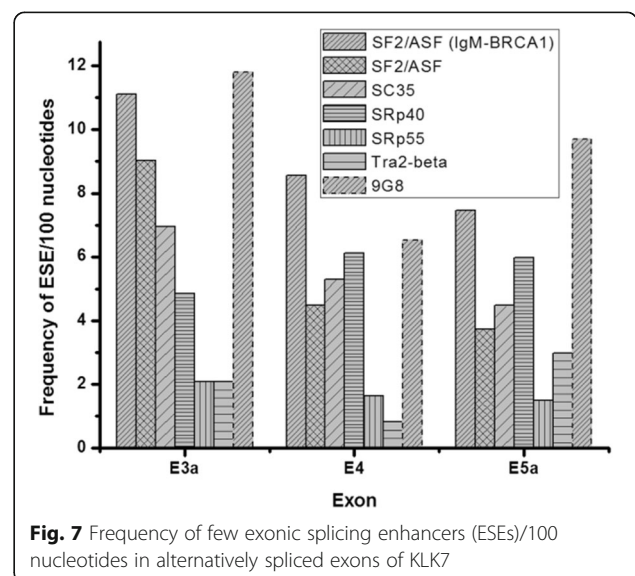
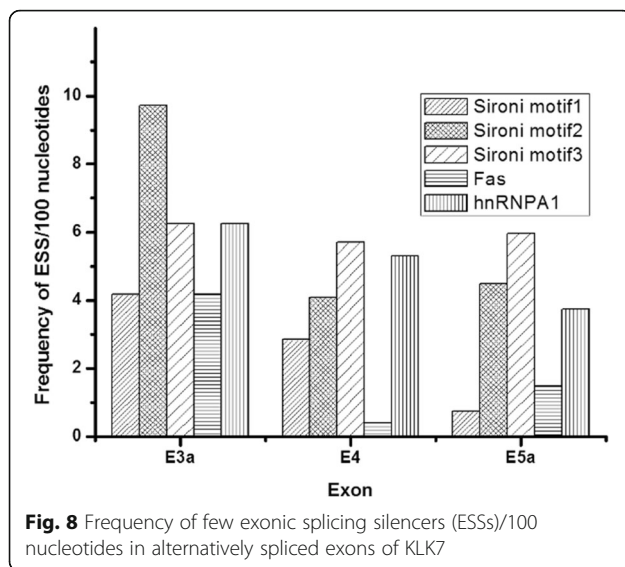


Fig. 7 Frequency of few exonic splicing enhancers (ESEs)/100 nucleotides in alternatively spliced exons of KLK7



benign breast disease patients). Hence it does not seem to be appropriate to relate down-regulated expression of KLK7 mRNA with various clinicopathological features. Similarly KLK7 mRNA expression level was not documented in normal breast tissues because none of the normal control consented to donate normal tissue. Keeping in view the fact that the tissue collected from patient apparently away from problematic region doesn't ensure absence of abnormality.

Followed by the study presented in this manuscript to explore and ascertain the outcome of KLK7 mRNA's down-regulated expression we quantified KLK7 protein (hK7) in the sera of breast cancer, benign breast disease patients and normal controls. Comparison indicated that hK7 level significantly goes down in breast cancer and benign breast disease patients as compared to normal controls. Statistical significance level (p -value) was calculated and correlation of hK7 down-regulation with various clinicopathological features was determined. Data has been accepted for publication somewhere else.

Although, the protein encoded by unexpected splice variant SER4/KF963192 has considerable homology with the reference protein, it will be a truncated protein. Moreover, due to the exclusion of E4 and truncation of E3 at 3' end, encoded protein molecule seems to be lacking the conventional aspartic acid (encoded by E4) and histidine residue (encoded by 3' end of E3) of histidine: aspartic acid: serine catalytic triad which is the characteristic feature of serine proteases. 3D structure prediction has further confirmed that the possible 3D structure of encoded protein molecule resembles with plasminogen (Fig. 5) rather than KLK7 protein (i.e., hK7). SNPs observed in breast cancer patients of our study are entirely different from the previously reported KLK7 transcripts and are being reported for the first time.

Using nested set of primers F1, F2, R1 & R2 five potential candidate transcripts lacking either exon 1 or 2 (Fig. 1) were missed. We also tried using primers targeting other exonic regions like 2, 3 & 4 but could not observe any positive PCR results (data not shown in the manuscript). Integrity of RNA was ensured through amplification of positive control, β -actin.

It is possible that genetic mutations might have resulted in the creation of new and stronger splice sites at the position of E3b and E5b and hence these splice sites would have been preferred. Another reason can be the involvement of ESEs/ISEs, ESSs/ISSs and trans-acting proteins which are known to favor selection of one splice site over other and regulate splicing pattern [26–28]. To better understand the mechanism underlying aberrant expression and alternative splicing pattern of KLK7 sequence of KLK7 gene must be known. Hence, it is essential to isolate and sequence KLK7 gene of breast cancer tissues used in our study. The predicted pre-mRNAs were developed by combining exonic sequences determined during the present study and intronic sequences already reported in the literature. We hypothesize that KLK7 genes in our study subjects might have variations in the intronic sequences that may influence splicing pattern. Hence it is not possible to justify the splicing event of SER4 solely on the basis of the distribution frequency of splicing enhancer and silencer motifs determined in the present study. Moreover, distribution of the trans-acting factor also contributes towards the regulation of splicing pattern and we are not aware about distribution of trans-acting proteins in breast tissues of our study. Further work is in progress in our lab regarding mechanistic details of the splicing mechanism.

Conclusions

To the best of our knowledge this study is the first to report expression of a new alternatively spliced variant of KLK7 (SER4; **KF963192**) in 02 patients, one suffering from mammary dysplasia and other suffering from invasive ductal carcinoma grade II. Alternatively spliced variant is a novel variant because it is generated by the combined effect of three splicing events including exclusion of E4 and truncations of the flanking exons E3 and E5 at 3' and 5' ends, respectively. Simultaneous exclusion of multiple adjacent exons has already been reported in mouse intercellular cell adhesion molecule 1 (ICAM-1) gene [29] but the splicing event that has been observed in case of KLK7 mRNA4 is unique because such splicing scheme has not been reported earlier. It may be population specific or disease specific variant but at present data is not sufficient to conclude about this aspect. There is need to further explore the role of this unexpected splice variant in pathobiology of mammary dysplasia.

KLK7 expression is down-regulated in most of the cancerous and benign breast disease patients. KLK7 isoforms or variants cannot be significantly correlated with breast cancer because KLK7 mRNA expression could only be detected in 5% of analyzed samples. According to data presented in Table 5 and Table 8, only 6 breast tissues (3 normal, 2 cancerous and 1 benign) were found positive for KLK7 expression. This observation implies the role of some epigenetic mechanism or transcriptional level control that maintains down-regulation of KLK7. There is need to explore underlying mechanism and identify factors that contribute towards down-regulation of KLK7 in breast cancer and benign breast disease patients.

Additional file

Additional file 1: Table S1. Predicted sequences of proteins encoded by constitutively spliced KLK7 mRNAs expressed in breast cancer tissues. **Table S2.** Probable sub-cellular localization of the predicted proteins encoded by KLK7 mRNAs in patients. **Table S3.** Splice site sequences of KLK7 transcripts. **Table S4.** Splice site scores and confidence level of KLK7 mRNAs splice sites. (DOCX 59 kb)

Abbreviations

cDNA: Complementary DNA; E: Exonic sequence; hK7: Human tissue Kallikrein 7 protein; hSCCE: Human stratum corneum chymotryptic enzyme; KLK: Kallikrein; KLK7: Human tissue Kallikrein 7 gene; pI: Isoelectric point; SER: Samina Ejaz RNA; SNP: Single Nucleotide Polymorphism; TD PCR: Touch down polymerase chain reaction

Acknowledgements

We are thankful to patients participating in this study.

Availability of data and materials

All data and material related with study is included in this manuscript and Additional file 1.

Funding

Financial support for study to Samina Ejaz, recipient of a Higher Education Commission (HEC), Pakistan scholarship, is hereby acknowledged.

Authors' contributions

FHN conceived this study, participated in the design and revised the manuscript. SE carried out all the bench work in Pakistan and wrote the first draft. MA was involved in bioinformatics analysis of the data and helped to revise the manuscript. GA being clinician helped in clinical diagnosis and collection of breast cancer tissue samples. All authors reviewed and approved the final manuscript.

Ethics approval and consent to participate

The study was approved by appropriate institutional technical committees including the departmental Ethics and Research Committee and by the Advanced Studies and Research Board (ASRB) of the Islamia University of Bahawalpur, Pakistan. Written Consent to participate in the study was taken on a prescribed form from all patients included in the study. The experimental work involving humans was in complete compliance with the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

It is certified that none of the authors have any conflict of interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Biochemistry and Biotechnology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. ²Department of Chemistry, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. ³Department of Surgery, Quaid-i-Azam Medical College, Bahawal Victoria Hospital, Bahawalpur, Pakistan.

Received: 2 March 2017 Accepted: 22 August 2017

Published online: 19 October 2017

References

- Lareau LF, Green RE, Bhatnagar RS, Brenner SE. The evolving roles of alternative splicing. *Curr Opin Struct Biol.* 2004;14(3):273–82.
- Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, et al. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *PNAS.* 2006;103(22):8390–5.
- David A, Majeesh N, Azar I, Biton S, Engel S, Bernstein J, et al. Unusual alternative splicing within the human Kallikrein genes KLK2 and KLK3 gives rise to novel prostate-specific proteins. *J Biol Chem.* 2002;277(20):18084–90.
- Brinkman BM. Splice variants as cancer biomarkers. *Clin Biochem.* 2004;37(7):584–94.
- Omenn GS, Yocum AK, Menon R. Alternative splice variants, a new class of protein cancer biomarker candidates: findings in pancreatic cancer and breast cancer with systems biology implications. *Dis Markers.* 2010;28(4):241–51.
- Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene.* 2016;35:2413–27.
- Michael IP, Sotiropoulou G, Pampalakis G, Magklara A, Ghosh M, Wasney G, et al. Biochemical and enzymatic characterization of human kallikrein 5 (hK5), a novel serine protease potentially involved in cancer progression. *J Biol Chem.* 2005;280(15):14628–35.
- Pampalakis G, Sotiropoulou G. Tissue kallikrein proteolytic cascade pathways in normal physiology and cancer. *Biochim Biophys Acta.* 2007;1776(1):22–31.
- Tan OL, Whitbread AK, Clements JA, Dong Y. Kallikrein-related peptidase (KLK) family mRNA variants and protein isoforms in hormone-related cancers: do they have a function? *Biol Chem.* 2006;387(6):697–705.
- Michael IP, Kurlender L, Memari N, Yousef GM, Du D, Grass L, et al. Intron retention: a common splicing event within the human kallikrein gene family. *Clin Chem.* 2005;51(3):506–15.
- Yousef GM, Scorilas A, Jung K, Ashworth LK, Diamandis EP. Molecular cloning of the human kallikrein 15 gene (KLK15). Up-regulation in prostate cancer. *J Biol Chem.* 2001;276(1):53–61.
- Kurlender L, Borgono C, Michael IP, Obiezu C, Elliott MB, Yousef GM, et al. A survey of alternative transcripts of human tissue kallikrein genes. *Biochim Biophys Acta.* 2005;1755(1):1–14.
- Hansson L, Stromqvist M, Backman A, Wallbrandt P, Carlstein A, Egelrud T. Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase. *J Biol Chem.* 1994;269(30):19420–6.
- Yousef GM, Scorilas A, Magklara A, Soosaipillai A, Diamandis EP. The KLK7 (PRSS6) gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family - genomic characterization, mapping, tissue expression and hormonal regulation. *Gene.* 2000;254(1–2):119–28.
- Dong Y, Kaushal A, Brattsand M, Nicklin J, Clements JA. Differential splicing of KLK5 and KLK7 in epithelial ovarian cancer produces novel variants with potential as cancer biomarkers. *Clin Cancer Res.* 2003;9(5):1710–20.
- Holzschelter L, Biermann JC, Kotsch M, Prezas P, Farthmann J, Baretton G, et al. Quantitative reverse transcription-PCR assay for detection of mRNA encoding full-length human tissue kallikrein 7: prognostic relevance of KLK7 mRNA expression in breast cancer. *Clin Chem.* 2006;52(6):1070–9.
- Talieri M, Diamandis EP, Gourgiotis D, Mathioudaki K, Scorilas A. Expression analysis of the human kallikrein 7 (KLK7) in breast tumors: a new potential biomarker for prognosis of breast carcinoma. *Thromb Haemost.* 2004;91(1):180–6. doi:10.1267/THRO0401018004010180.

18. IARC. Press release, Latest world cancer statistics. www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf. 2013.
19. Junaidi I. Breast cancer claims 40,000 lives a year in Pakistan. Dawn. 2012 10th January; p <http://dawn.com/2012/01/10/breast-cancer-claims-400000-lives-a-year-in-pakistan-2/>.
20. Saleem S. Death from breast cancer more likely for Pakistani patients: US expert. The Express Tribune. 2011 22nd January, 2011; p. tribune.com.pk/story/107232/death-from-breast-cancer-more-likely-for-pakistani-patients-us-expert/.
21. Bhurgri Y, Pervez S, Kayani N, Bhurgri A, Usman A, Bashir I, et al. Cancer profile of Larkana, Pakistan (2000-2002). *Asian Pac J Cancer Prev*. 2006;7(4):518–21.
22. Sohail S, Alam SN. Breast cancer in pakistan - awareness and early detection. *Journal of College of Physicians and Surgeons Pakistan*. 2007;17(12):711–2.
23. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc*. 2008;3(9):1452–6. doi:10.1038/nprot.2008.133.
24. Lindqvist R. Detection of *Shigella* spp. in food with a nested PCR method-sensitivity and performance compared with a conventional culture method. *J Appl Microbiol*. 1999;86(6):971–8.
25. Marsilio F, Di Martino B, Decaro N, Buonavoglia C. A novel nested PCR for the diagnosis of calicivirus infections in the cat. *Vet Microbiol*. 2005;105(1):1–7. doi:10.1016/j.vetmic.2004.09.017.
26. Blencowe BJ. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci*. 2000;25(3):106–10. doi:10.1016/S0968-0004(00)01549-8.
27. Havlioglu N, Wang J, Fushimi K, Vibranovski MD, Kan Z, Gish W, et al. An intronic signal for alternative splicing in the human genome. *PLoS One*. 2007;2(11):e1246. doi:10.1371/journal.pone.0001246.
28. Wang J, Smith PJ, Krainer AR, Zhang MQ. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res*. 2005;33(16):5053–62. doi:10.1093/nar/gki810.
29. Mizgerd JP, Spieker MR, Lupa MM. Exon truncation by alternative splicing of murine ICAM-1. *Physiol Genomics*. 2002;12(1):47–51. doi:10.1152/physiolgenomics.00073.2002.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

