Investigating isoform switching in RHBDF2 and its role in neoplastic growth in breast cancer (#69682)

First submission

Guidance from your Editor

Please submit by **20 Feb 2022** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

1 Latex file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Investigating isoform switching in RHBDF2 and its role in neoplastic growth in breast cancer

Mehar Masood [1], **Noor Us Subah** [1], **Maria Shabbir** [2], **Rehan Z Paracha** [1], **Mehak Rafiq** [Corresp. 1]

[1] Research Centre for Modelling and Simulation, National University of Sciences and Technology, Islamabad, Pakistan

[2] Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, Islamabad, Pakistan

Corresponding Author: Mehak Rafiq
Email address: mehak@rcms.nust.edu.pk

Breast cancer is the second leading cause of cancer-related deaths globally, and its prevalence rates are increasing day by day. One major reason is the assumption that a "one gene leads to one protein" approach in most studies undergoing predicting therapeutic drug targets for cancer therapy. Therefore, there is always an immense need to find promising and novel anti-cancer drug targets. Proteases have an integral role in cell proliferation and growth because the proteolysis mechanism is an irreversible process that aids in regulating cellular growth during tumorigenesis. Therefore, they can be considered an important target for cancer treatment. Apart classifying rhomboids into active and inactive rhom- boids, alternative splicing tends to produce more transcript isoforms of inactive rhomboids "iRhom2". Speculatively, previous studies on gene expression analysis of RHBDF2 gene encoding iRhom2 showed heterogenous behaviour during tumorigenesis. Consistent with this, several studies have reported the antagonistic role of iRhom2 in tumorigenesis, i.e. either they are involved in negative regulation of EGFR ligands via ERAD pathway or positively regulate EGFR ligands via EGFR signalling pathway. Additionally, different opinions suggest iRhom2 mediated cleavage of EGFR ligands takes place TACE dependently or TACE independently. However, how to reconcile these seemingly opposing roles is still unclear and might be attributed to more than one transcript isoform of iRhom2. To observe the differences at isoform resolution, the current strategy identified isoform switching in RHBDF2 via differential transcript usage using RNA-seq data during breast cancer initiation and progression. Furthermore, interacting partners were found via correlation and enriched to explain its antagonistic role. Isoform switching was observed at DCIS, grade 2 and grade 3 from canonical to the cub isoform. Neither EGFR nor ERAD was found enriched. However, pathways leading to TACE dependent EGFR signalling pathways were more observant, specifically MAPK signalling pathways, GPCR signalling pathways, and Toll-like receptor pathways. Nevertheless, it was noteworthy that during CTCs, the cub isoform switches back to the canonical isoform, the proteasomal

degradation pathway and cytoplasmic ribosomal protein pathways were found significantly enriched. Therefore, it could be inferred that cub isoform functions during cancer initiation in EGFR signalling whereas during metastasis where invasion is the primary task, isoform switches back to the canonical.

Investigating Isoform Switching in *RHBDF2* for its role in Human Neoplastic Growth in Breast

Mehar Masood[1], Noor us Subah[1], Maria Shabbir[2], Rehan Zafar Paracha[1], and Mehak Rafiq[1]

[1]**Research Centre for Modelling and Simulation, National University of Sciences and Technology, Pakistan**
[2]**Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, Pakistan**

Corresponding author:
Mehak Rafiq[1]

Email address: mehak@rcms.nust.edu.pk

ABSTRACT

Breast cancer is the second leading cause of cancer-related deaths globally, and its prevalence rates are increasing day by day. One major reason is the assumption that a "one gene leads to one protein" approach in most studies undergoing predicting therapeutic drug targets for cancer therapy. Therefore, there is always an immense need to find promising and novel anti-cancer drug targets. Proteases have an integral role in cell proliferation and growth because the proteolysis mechanism is an irreversible process that aids in regulating cellular growth during tumorigenesis. Therefore, they can be considered an important target for cancer treatment. Apart classifying rhomboids into active and inactive rhomboids, alternative splicing tends to produce more transcript isoforms of inactive rhomboids "iRhom2". Speculatively, previous studies on gene expression analysis of RHBDF2 gene encoding iRhom2 showed heterogenous behaviour during tumorigenesis. Consistent with this, several studies have reported the antagonistic role of iRhom2 in tumorigenesis, i.e. either they are involved in negative regulation of EGFR ligands via ERAD pathway or positively regulate EGFR ligands via EGFR signalling pathway. Additionally, different opinions suggest iRhom2 mediated cleavage of EGFR ligands takes place TACE dependently or TACE independently. However, how to reconcile these seemingly opposing roles is still unclear and might be attributed to more than one transcript isoform of iRhom2. To observe the differences at isoform resolution, the current strategy identified isoform switching in RHBDF2 via differential transcript usage using RNA-seq data during breast cancer initiation and progression. Furthermore, interacting partners were found via correlation and enriched to explain its antagonistic role. Isoform switching was observed at DCIS, grade 2 and grade 3 from canonical to the cub isoform. Neither EGFR nor ERAD was found enriched. However, pathways leading to TACE dependent EGFR signalling pathways were more observant, specifically MAPK signalling pathways, GPCR signalling pathways, and Toll-like receptor pathways. Nevertheless, it was noteworthy that during CTCs, the cub isoform switches back to the canonical isoform, the proteasomal degradation pathway and cytoplasmic ribosomal protein pathways were found significantly enriched. Therefore, it could be inferred that cub isoform functions during cancer initiation in EGFR signalling whereas during metastasis where invasion is the primary task, isoform switches back to the canonical.

INTRODUCTION

Breast Cancer is the second leading cause of cancer-related deaths globally, and its prevalence rates are increasing day by day. Cancer treatment has emerged as a serious challenge due to the inadequate availability of therapeutic targets (Mansoori et al., 2017) .One of the major reasons lies in stressing over the assumption of a "one gene to one protein to one functional pathway" approach in most studies undergoing predicting therapeutic drug targets for cancer therapy. Therefore, there is always an immense need to find promising and novel anti-cancer therapeutic drug targets. Proteases have an integral role in

cell proliferation and growth because the proteolysis mechanism is an irreversible process that aids in regulating cellular growth during tumourigenesis (Park et al., 2020). Therefore, they can be considered an important target for cancer treatment.

Rhomboid proteases hydrolyse the peptide bonds in other proteins and are almost found in all kingdoms of life. However, some members of the rhomboid family lack essential catalytic residues necessary for proteolysis, suggesting that they cannot cleave substrates. Rather, they can do so by complex formation with client proteins known as inactive rhomboids or pseudoprotease. One of the member's knowns as iRhom2 encodes by the gene RHBDF2 has developed a new pseudo enzyme function regulating trafficking, orchestrating inflammatory response and growth factor signalling by interacting with client proteins (Bergbold and Lemberg, 2013). Both active and inactive rhomboids have many transcript isoforms in mammals, with several of them can code for alternative forms of the proteins. Whereas iRhom2, have two functionally important isoforms, ENST00000313080 (canonical) and ENST00000591885 (cub) which are also reported in public databases (ENSEMBL and Refseq) along with 18 computationally mapped transcript isoforms.

It is anticipated that alternative splicing and its related proteins are thought to be involved in the dynamic phenotypic changes in cancer cells (Chabot and Shkreta, 2016). The importance of analysing isoforms instead of genes has been highlighted because cancer cell growth is directly linked to the aberrant use of one alternatively spliced formed isoform over another under unfavourable circumstances. One splicing phenotype in cancers can be Isoform switching (Soneson et al., 2020) .In mammalian cells, the dynamic phenotypic changes are mostly due to the growth, proliferation and differentiation held by EGFR signalling pathways to maintain homeostasis. However, epithelial cancers are characterized by enhanced EGFR activation (Nicholson et al., 2001). Aberrant functioning of EGFR could be due to overexpression of EGFR and its ligands, excessive transactivation of EGFR ligands, aberrant ligand processing or mutations affecting EGFR (Nickel et al., 1983). Important EGFR ligands such as AREG, HB-EGF, TGF-a and EGF binds to EGF receptors (EGFR's) as proproteins (inactive form) and must be cleaved to shed into the extracellular compartment inactive form for further signalling. These ligands are membrane-tethered and different proteases like pseudoproteases helps in cleaving membrane-bound EGFR pro-ligands to convert them into biologically active proteins during proliferation (Dulloo et al., 2019).

Several studies have reported the antagonistic role of iRhom2 , i.e. either they are involved in negative regulation of EGFR ligands via ERAD pathway or positively regulate EGFR ligands via EGFR signalling pathway. Additionally, different opinions suggest iRhom2 mediated cleavage of EGFR ligands takes place TACE dependently or TACE independently. Consistent with this, insilico analysis of publically available gene expression datasets on Breast Cancer showed heterogenous expression behaviour of RHBDF2 according to the intrinsic molecular subtypes and histological grading and staging. However, progression to carcinoma is not as simple and we found little literature for studying mRNA expression during neoplastic growth i.e, cancer initiation and progression. ICD-10 classifies neoplasms into four groups, benign neoplasm, in situ neoplasm, malignant neoplasm and neoplasm of uncertain or unknown behavior. Studies state that alternative splicing mechanism is pathologically altered during neoplastic growth impacting the cell behavior and causing tissue specific changes.

Proposed Work

In the present study, we explore the antagonistic role and heterogenous expression behaviour of RHBDF2 encoding iRhom2 during the neoplastic growth in Breast, considering the bidirectional role that might be attributed to the presence of more than one functionally important isoforms of RHBDF2 in mammals during tumorigenesis.

METHODS

Data Collection

The paired end fastq files containing raw reads for samples in each dataset (GSE52194,GSE130660, GSE69240, GSE110114,GSE45419,GSE51124,GSE148991) were downloaded using an FTP link from EMBL-EBI (The European Bioinformatics Institute), having accession numbers followed by the SRR acronym. The datasets were classified as follows (GSE52194, GSE130660 for normal versus primary tumour), (GSE69240 for normal versus ductal carcinoma insitu- DCIS), (GSE110114, GSE45419 for

normal versus invasive ductal carcinoma- IDC), (GSE148991 for normal versus circulating tumour cells-CTCs) and (GSE51124 for normal versus grade2 and grade3).

Transcriptome Reconstruction and Quantification

The raw data was pre-processed for adapter sequences using fastp (Soneson et al., 2020). The transcriptome data was then analyzed using new Tuxedo pipeline (Pertea et al., 2016). The filtered reads were aligned on the reference genome GRch38 using Hisat2. The mapped reads from each sample along with the genome GTF file were used to perform annotation-based transcriptome assembly using StringTie. The assemblies were then compared and merged. The StringTie merge function creates a set of merged transcripts that are comparable for the subsequent analysis.

Differential Isoform Expression

StringTie produces a set of read or coverage tables/files of the quantification or abundance data that were read into R for isoform expression analysis using IsoformSwitchAnalyzeR. Importing data include preparing a transcript sequence FASTA file, a parent directory containing coverage table/files, quantification files of the samples in GTF format, and a design file enlisting the phenotypic data, i.e. sample ID and its corresponding condition. The FASTA sequence file for transcripts was generated using a program utility called gffread (Pertea and Pertea, 2020). The utility generates a FASTA file with DNA sequences for all the transcripts present in the GTF file.The inter-sample normalisation was done using edgeR embedded in the importRdata() function that concatenates all the information into SwitchAnalyzeRlist. This SwitchAnalyzeRlist acts as an object containing all the data frames and phenotypic data related to the dataset. The abundance files generated via StringTie are already normalised for intra-sampling using the FPKM approach. EdgeR works best for inter-sample normalisation via the TMM (trimmed mean of M values) method on pre-normalised count data (Maza, 2016).The normalised data was prefiltered to remove the uninterested data of transcripts and genes from the switch list object such as non-expressed isoforms or genes and genes with only one isoform. The differential isoform usage test was performed using the function isoformswitchTestDEXSeq() enabling the switch identification.

Annotating Unknown transcript isoform

During stringTie-merge step, transcripts are labelled as MSTRGs. These could sometimes be either novel transcripts, false positives or true transcripts that are left unannotated.The transcripts with MSTRGs labels were annotated via BLASTp. To ensure either they are novel or already exist and are mislabelled. BLASTp check the sequence similarity of these MSTRGs to already annotated transcript sequences deposited in the public databases. The input sequence of these MSTRGs for BLASTp was extracted via extract sequence function(). The MSTRGs were further considered for downstream analysis based on E-value cutoff = 0, per cent identity .ie. 100% and query length matching the length of the already existing annotated transcript i.e, the target sequence length. For canonical transcript (ENST00000313080) it should be 856 aa and for cub transcript (ENST000591885) it should be 827 aa. The final transcripts were analysed to predict their functional consequences using external tools. CPC2 (jian Kang et al., 2017) was used to check the coding potential and pFam (Finn et al., 2016) to predict the biological domains. The input FASTA files for the tools were manipulated using the Seqkit (Shen et al., 2016) package according to the requirements of each tool.

Finding interacting partners via Correlation

The interacting partners were found using correlation. prepDE.py python script was used to obtain read count information from the quantification file generated via StringTie. The count files were then subjected for differential analysis on the GALAXY server using DESeq2.The normalised differential counts were then input for correlation analysis at significance level $p-value < 0.05$ and correlation cutoff ± 0.7.

Enrichment

The most significant correlated partners were enriched using GSEA (Gene set enrichment analysis) (Subramanian et al., 2005).

RESULTS

Annotation of MSTRGs via BLASTp

To ensure either MSTRGs labelled transcripts are novel or already exist and are mislabelled, BLASTp was used to annotate them (Acland et al., 2013). If any of these transcripts would match the already annotated transcripts based on the query length, percent identity, and E-value as already mentioned, they would be merged with already annotated transcripts for further analysis.

The MSTRG labelled transcripts (MSTRG.15617.1, MSTRG.15617.2, MSTRG.15617.3) from the dataset GSE110114 of query length 827 amino acid (AA) showed 100% identity with isoform-2 of RHBDF2 under accession id (NP_001005498.2) of length 827 AA exactly matching the query length, i.e. 827 AA and E-value 0. Similarly, following MSTRG labelled transcripts fullfilled the above mentioned criteria : the transcripts (MSTRG.20860.1, MSTRG.20860.2, MSTRG.20860.3, MSTRG.20860.4) from the dataset GSE69240, the transcripts (MSTRG.17267.4 and MSTRG.17267.5) from the dataset GSE52194, the transcripts (MSTRG.17431.2, MSTRG.17431.3, MSTRG.17431.4, MSTRG.17431.5, MSTRG.17431.8, MSTRG.17431.9, MSTRG.17431.10) from the dataset GSE45419, the transcript (MSTRG.15555.1) from the dataset GSE130660, the transcripts (MSTRG.19489.1,MSTRG.19489.3, MSTRG.19489.4, MSTRG.19489.5, MSTRG.19489.8) from the dataset GSE51124, the transcripts (MSTRG.22199.1, MSTRG.22199.5, MSTRG.22199.7, MSTRG.22199.9) from the dataset GSE148991. These transcripts were further selected for analysis also shown in Table 1.

From the dataset GSE52194, the transcript (MSTRG.17267.1), from the dataset GSE45419, the transcripts (MSTRG.17431.1, MSTRG.17431.13), from the dataset GSE148991, the transcripts (MSTRG.22199.2, MSTRG.22199.10) of query length 856 AA showed 100% identity with isoform-1 of RHBDF2 under accession id (NP_078875.4) amino acid length 856 AA exactly matching the query length, i.e. 856 AA and E-value 0. These transcripts were further selected for analysis also shown in Table 2.

Dataset	Similarity match with cub isoform
GSE52194	MSTRG.17267.4, MSTRG.17267.5
GSE130660	MSTRG.15555.1
GSE69240	MSTRG.20860.1, MSTRG.20860.2, MSTRG.20860.3, MSTRG.20860.4
GSE110114	MSTRG.15617.1,MSTRG.15617.2,MSTRG.15617.3
GSE45419	MSTRG.17431.2, MSTRG.17431.3, MSTRG.17431.4, MSTRG.17431.5, MSTRG.17431.8 MSTRG.17431.9, MSTRG.17431.10
GSE51124	MSTRG.19489.1, MSTRG.19489.3, MSTRG.19489.4, MSTRG.19489.5,MSTRG.19489.8
GSE148991	MSTRG.22199.1, MSTRG.22199.5, MSTRG.22199.7, MSTRG.22199.9

Table 1. MSTRG labelled transcripts showing similarity match with cub isoform.

Dataset	Similarity match with canonical isoform
GSE52194	MSTRG.17267.1
GSE45419	MSTRG.17431.1, MSTRG.17431.13
GSE148991	MSTRG.22199.2, MSTRG.22199.10

Table 2. MSTRG labelled transcripts showing similarity match with canonical isoform.

Cross-validation of BLASTp hit sequences

The MSTRGs transcripts showing 100% identity and similarity match hit with NCBI sequences using BLASTp. The NCBI protein sequence database sources are RefSeq and Genbank mainly. ENSEMBL gene sets are derived from multiple sources, partly from RefSeq, partly from uniport and partly from Havana annotation. Although, as stated above, the criteria have been set for matching query length using ENSEMBL annotated transcripts, it was important to find whether a transcript from ENSEMBL has a close match to the transcript from BLASTp hit. To cross-validate, the ENSEMBL annotated sequences

to the corresponding BLASTp annotated sequences, multiple sequence alignment was performed via CLUSTAL. The FASTA sequences were extracted from the NCBI for accession id (NP_001005498.2 and NP_078875.4) corresponding to selected MSTRG transcripts. The alignment showed 100% similarity between NP_001005498.2 and ENST0000591885 (cub) and 100% similarity between NP_078875.4 and ENST0000313080 (canonical). Thus, selected MSTRGs were merged with already annotated transcripts and were further subjected for calculations.

Isoform Expression

The Figure 1 shows the isoform expression via parameter isoform fraction of the two transcript isoforms canonical and the cub across the two conditions. Isoform fraction is calculated by taking the ratio between isoform expression and gene expression values in each condition. The isoform expression in normal versus primary tumor conditions (GSE52194 and GSE130660) of both the transcripts (canonical and cub) significantly decreases. Whereas, for the dataset(GSE110114 and GSE45419) where the comparison is between normal versus IDC condition the change in the isoform expression is insignificant. Interestingly for the dataset (GSE69240) normal versus DCIS, (GSE51124) normal versus grade 2 and grade 3, (GSE148991) normal versus CTCs , the change in isoform expression is statistically significant.

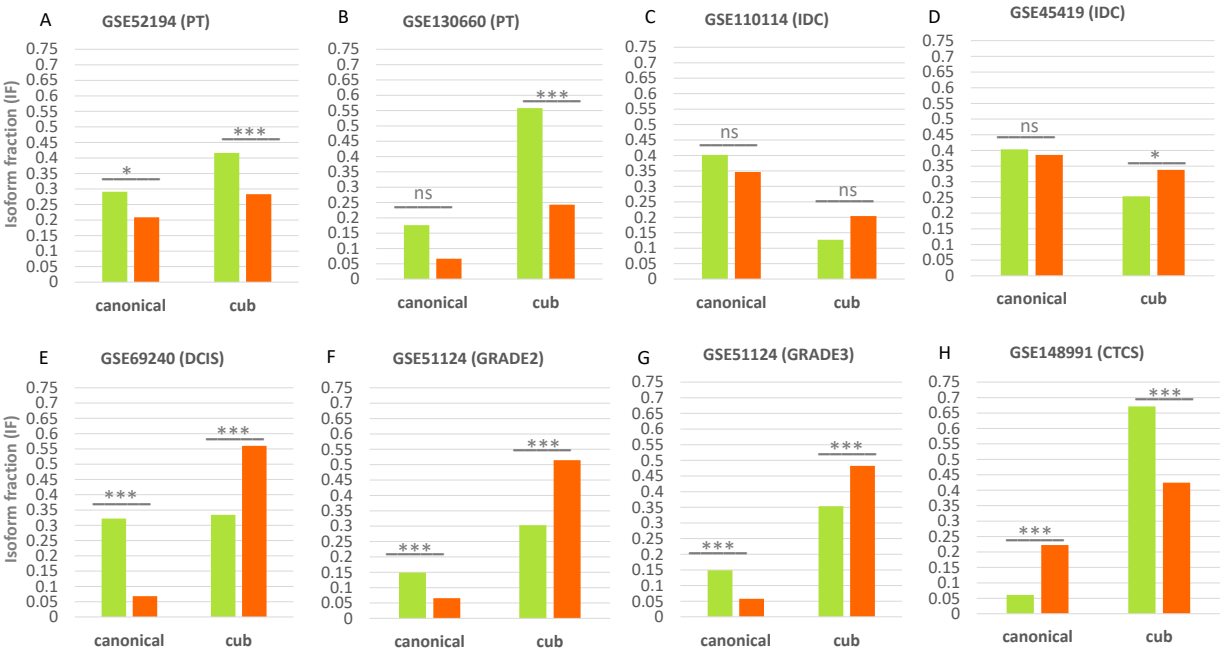

Figure 1. Isoform fraction of canonical and cub transcript in normal and tumour condition. **A-H**, The primary tumour dataset shows that isoform fraction of both the isoforms canonical and cub are significantly less in tumour state whereas in IDC the change in isoform fraction is insignificant. For DCIS, grade2 and grade3 isoform fraction of cub isoform in tumour state is significantly more than canonical isoform. Interestingly, in CTCs cub isoform fraction start decreasing significantly.

Differential Isoform Usage

Two parameters are considered for significant isoform switching i.e, the statistical significance and the effect size. Statistical significance is calculated via $p-value$ and it should be <0.05. Whereas effect size tells the association between the two variables and here Differential isoform fraction (dIF) measures the change in the isoform fractions of the pair of isoforms across the condition. For isoform switching, in each dataset, the pair of isoforms should show opposite increase or decrease in the isoform usage across the conditions. It is calculated by taking the difference between the isoform fraction values. The cut off

for dIF ≥ 0.05. The Figure 2 shows that the isoform switches from canonical to the cub transcript isoform at DCIS, grade2 and grade3 whereas, it switches back to the canonical isoform at CTCs.

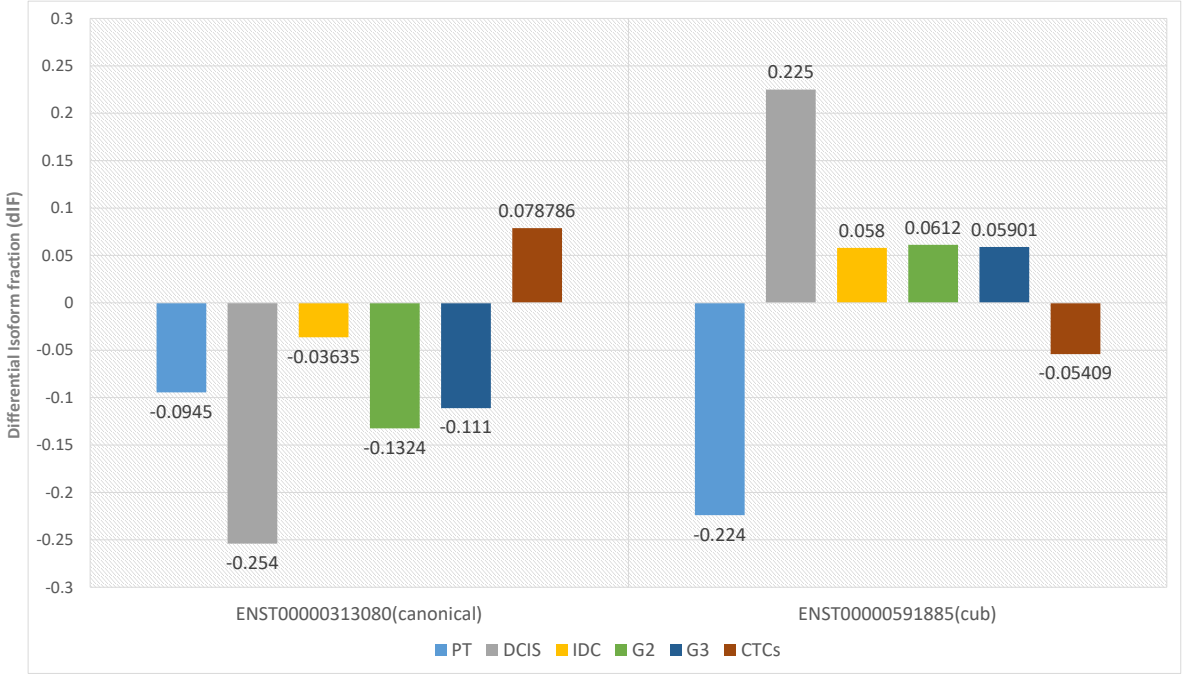

Figure 2. Differential isoform fraction of canonical and cub transcript. The blue bar represent that both the isoforms canonical and cub showed decreased usage. During DCIS (gray), grade2 (green), grade3 (dark blue) canonical isoform shows decreased usage whereas cub isoform shows increased usage. During IDC (yellow) although there is an opposite usage but unable to fulfill dIF cutoff. Interestingly, during CTCs (brown bar) isoform expression is revereted where cub isoform shows decreased usage and canonical isoform shows incresaed usage.

Relative isoform fraction

The analyses were done on two primary tumour datasets and two IDC datasets, hence to make better inference, the IF values of the two primary tumour datasets and two IDC datasets were merged by taking the average values, even though a better approach would have been to do the meta-analysis but it has been seen in the literature that this approach also gives comparable results. The final plot in Figure 3 is constructed, showing the relative usage of the isoform fractions of the two isoforms across conditions. Here, in normal vs primary tumour, the relative change of the two isoforms across two conditions remains the same, i.e. in normal conditions, canonical is 30% and the cub is 70% similarly in tumour condition canonical is 30%, and the cub is 70%. However, in normal vs DCIS, it can be seen that in normal conditions, the contribution of the canonical and cub transcript to the overall gene expression is equal, i.e. 50%, while in tumour condition, cub transcript expression rises up to being 90% to the overall gene expression relative to the canonical transcript which is just 10%. Hence, it can be said that a switch in expression has occurred. Surprisingly, the same trend can be seen in normal vs IDC like normal vs primary tumour. The relative change of the two isoforms across two conditions is the same, i.e. in normal conditions, canonical is 68%, and the cub is 32%. Similarly, in tumour condition, canonical is 68%, and the cub is 32%. In normal vs grade 2, in normal condition, the relative usage of the canonical transcript is 22%, and cub transcript is 78%. In contrast, there is a drastic change in relative usage across the conditions. Here, the cub transcript contributes 94% to the overall gene expression compared to canonical, which is 6%. Similarly, in normal vs grade 3, in normal condition, the relative usage of the canonical transcript is 22%, and cub transcript is 78%. In comparison, in tumour conditions, the relative usage

of the canonical transcript is 8%, and the cub transcript is 92%. Finally, in normal vs CTCs, in normal conditions, the relative usage of the canonical transcript is 9%, and cub transcript is 91%. In comparison, in tumour conditions, the relative usage of the canonical transcript is 28%, and the cub transcript is 72%. Hence, it is clear from the plot that isoform switching from canonical to cub transcript exists at DCIS, grade 2 and grade 3, while at CTCs, the transcript again switches from cub to canonical.

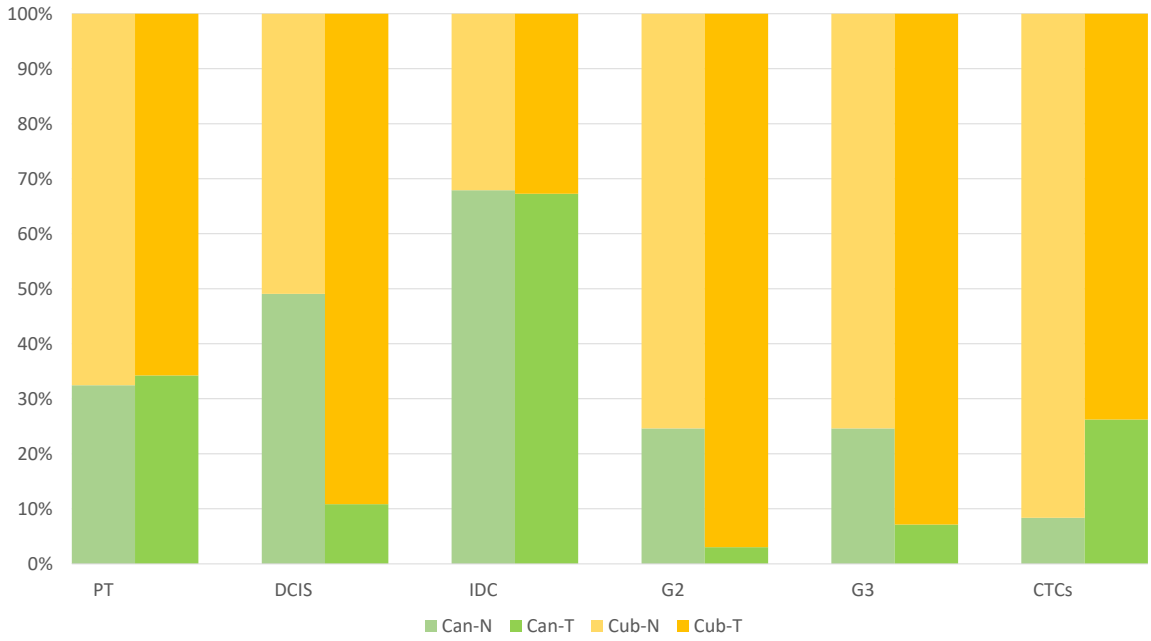

Figure 3. The plot shows the relative isoform usage of the canonical and cub isoform in normal and tumour conditions. The light green bar represents canonical isoform usage in normal conditions (Can-N), and the dark green bar represents canonical isoform usage in tumour conditions (Can-T). In contrast, the light orange bar represents cub isoform usage in normal condition (Cub-N) and the dark orange bar represents cub isoform usage in the tumour condition (Cub-T).

Finding Interactive partners via Correlation

Isoform switching was observed at DCIS, grade 2, grade 3 and CTCs, so it was important to find the interacting gene partners via correlation analysis. Correlation analysis at +0.7 to -0.7 cut-off to find the statistically significant correlated gene partners was performed. Interacting genes were then subjected for enrichment analysis via GSEA (Subramanian et al., 2005). GSEA determines a defined set of genes and their biologically meaningful interpretation across two conditions or phenotypes. Since our interacting genes were already ranked according to the correlation criteria, the analysis was done using pre-ranked GSEA. Pre-ranked GSEA was run on default parameters except for the minimum gene set size parameter set to 5, i.e. gene sets smaller than five were excluded from the analysis. There were no common interacting gene partners among DCIS, grade 2 and grade 3, so the union of interacting gene partners and their correlation values were input for pre-ranked GSEA as a ranked list of genes. Since during CTCs, the cub isoform switches back to canonical isoform in tumour condition, a list of the ranked gene for CTCs was run separately. The biological and molecular processes corresponding to mitogen-activated protein kinase (MAPK), G-protein coupled receptor (GPCR) and Toll-like receptor-related signalling pathways were most commonly observed among all the databases. Table 3 shows the GSEA results for the interacting partners for DCIS,G2 and G3 and Table 4 shows for CTCs along with p-value and the number of the genes in an enriched geneset.

Biological processes	Size	$p-Value$
MAPK SIGNALING PATHWAY	38	0
G PROTEIN SIGNALING PATHWAYS	12	0.008475
TOLLLIKE RECEPTOR SIGNALING PATHWAY	6	0.00352

Table 3. Enriched biological processes for the most correlated genes with RHBDF2 during DCIS, GRADE 2 and GRADE 3.

Biological processes	Size	$p-Value$
MAPK SIGNALING PATHWAY	16	0.047
G PROTEIN SIGNALING PATHWAYS	9	0.043
TOLLLIKE RECEPTOR SIGNALING PATHWAY	5	0.006
PROTEASOME DEGRADATION	7	0.0026
CYTOPLASMIC RIBOSOMAL PROTEINS	17	0.004

Table 4. Enriched biological processes for the most correlated genes with RHBDF2 during CTCs.

Running Leading Edge Analysis

Not all the members in a geneset are particularly contributing to the biological pathway. Often it is useful to extract the core genes that are contributing more to the enrichment score of the significant biological pathways.The leading-edge subset in a gene set is those genes that appear in the ranked list at or before the point at which enrichment score (ES) reaches its maximum deviation from zero. After running GSEA, leading-edge analysis helps to examine the genes in the leading-edge subsets of the enriched gene sets. A gene in many of the leading-edge subsets is more likely to be of interest than a gene in only a few of the leading-edge subsets. The subset of genes from the leading-edge analysis is shown in Figure 4. The heatmap shows the names of those subsets of genes that were found mostly enriched in in one or all genesets. RELA, RELB, RRAS, GNA12, PRKACA are the interesting genes found in 2 out of 3 gene sets for DCIS, grade2 and grade3 whereas NFKB1 is the only interesting gene found in 2 out of 5 gene sets for CTCs. Here, genesets corresponds to the biological process that are enriched.

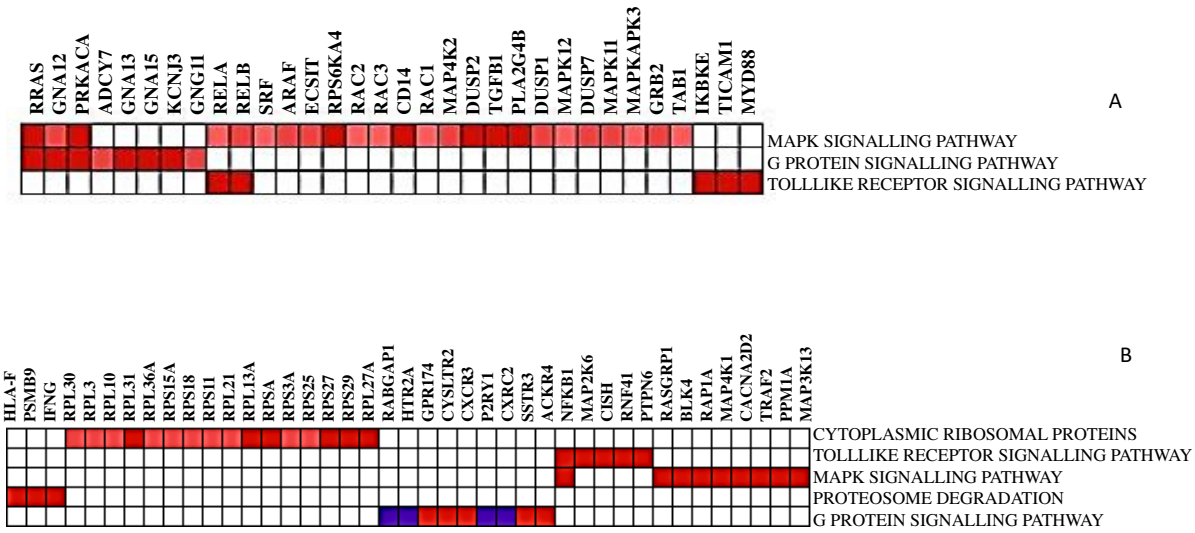

Figure 4. The heat map shows the (clustered) genes in the leading-edge subsets.**Heatmap(A)** is for DCIS, g rade2, and grade 3, whereas **Heatmap(B)** is for CTCs. The range of colours (red, pink, light blue, dark blue) shows the range of correlation values (high, moderate, low, lowest) in an enriched geneset.

DISCUSSION

Breast cancer has become the leading cause of oncologic mortality and morbidity among women worldwide. Virtually all breast carcinomas appear to originate from the uncontrolled production of epithelial cells of breast tissues forming a lump as shown in Figure 5. When normal epithelium begun to undergo malignant transition, the first progressive phase of excessive proliferation known as hyperplasia occurs, followed by the appearance of aberrant cells. At a later phase known as carcinoma in situ, these cells acquire malignant phenotype but lack invasive property due to the loss of cell motility. In the final phase of progression, the cell undergoes complete morphological changes causing the cells to break through basal membranes, thereby becoming invasive carcinoma (Allred et al., 2001). These cancer cells then grow into a solid tumour, eventually causing new blood vessels to grow, called angiogenesis. Next, they invade through EMT, a process called invasion, and get into the bloodstream (intravasation). Finally, these tumour cells are called CTCs (circulating tumour cells). They extravasate into the secondary site surviving the microenvironment. They grow into metastatic cancer or sometimes remain dormant for years (Harbeck et al., 2019).

Apart from the transcriptional factors, it is anticipated that alternative splicing factors and their related proteins are thought to be involved in regulating the cellular programs and the dynamic phenotypic changes in cancer cells (Chabot and Shkreta, 2016). Studies have shown that splicing is often pathologically altered, impacting cell behaviour during cancer initiation and progression (xiang Lu et al., 2015). More than 90% of the eukaryotic genes in mammals generate multiple isoforms, and aberrant splicing has become the cause of many diseases in humans (Sorek et al., 2004). The investigation of aberrant splicing events and isoform quantification has been shaped not only by the development of the techniques but also by the statistical approaches and algorithms being designed for their authentic biological interpretations. With the advent of the AS analysis tools, it became necessary to categorize how existing gene-level expression can be differentiated from transcript-level expression. The importance of analysing isoforms instead of genes has been highlighted by many examples showing functionally essential changes in the body (Zhang et al., 2013). Typically, coding genes have a transcript isoform expressed significantly

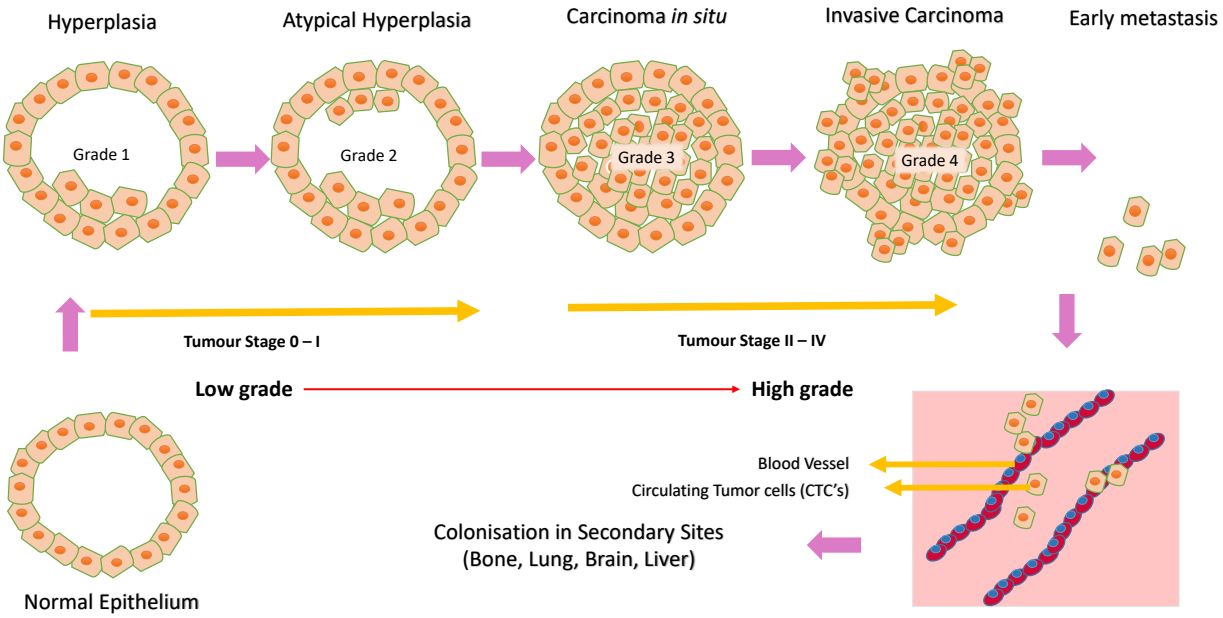

Figure 5. cancer initiation and progression.Steps in neoplastic growth during Breast Cancer

higher than other alternatively spliced transcript isoforms, often known as canonical isoforms. Under unfavourable circumstances like disease states, the dominance completely shifts from canonical to the other alternative transcript isoforms. One splicing phenotype in cancers can be Isoform switching (Di et al., 2018). Isoform switching is the relative abundance of different isoforms of the same gene that is reversed in different cell types or when disease tissue are compared to normal tissues.

Besides the classification into active rhomboids and inactive rhomboids, mechanistic alternative splicing produces even more transcript variants or isoforms with distinct functionalities. Interestingly, mammal harbours two inactive rhomboids, namely iRhom1 encoded by the RHBDF1 gene and iRhom2 encoded by the RHBDF2 gene. Both share highly conserved protein sequences, and the distinction lies in the protein sequences of the cytosolic region, where they possess different deletions and extensions. The KO studies on iRhom2 showed more severe phenotypic changes, whereas the phenotype of iRhom1 KO mice is much less clear. Thus, this makes iRhom2 an interesting gene to study.Several studies have reported antagonistic role of iRhom2 in tumorigenesis and other diseases i.e, either they are involved in negative regulation of EGFR ligands via ERAD pathway or they positively regulate EGFR ligands leading to EGFR signaling pathway. There are parallel studies suggesting iRhom mediated cleavage of EGFR ligands via TACE dependent or TACE independent pathway (Al-Salihi and Lang, 2020).

Some studies claim iRhom2 to be negative regulators of EGFR ligands (Lee et al., 2016). Evidence showed onset of sleep like phenotype in D.melanogaster is because of iRhoms are involved in negative regulation of EGFR signaling through ERAD pathway in nervous system whereas active rhomboids regulate cleavage of EGFR membrane bound precursors (Adrain and Freeman, 2012) .There exist some conserved mechanistic link between mammals and drosophila in regulation of EGFR signaling and in maintaining cell quality control machinery for efficient trafficking (Etheridge et al., 2013). iRhom2 can negatively regulate EGFR signaling via breakdown of EGF like substrates (Lee et al., 2016).They increase ERAD activity by bringing clients passively by delaying ER retention, hence enhancing the chance of exposure to ERAD machinery. While (Zettl et al., 2011) suggest that they can perform this mechanism by specifically destabilizing some substrates in ER, inhibiting their access to active rhomboids leading to degradation. Apart from cancer, (Lyu et al., 2018) identified the high expression of iRhom2 in renal tubules as target of PPARG thus promoting EGF degradation via ERAD.

Whereas, (Hosur et al., 2014) debate on EGFR signaling independent of TACE and states that cleavage of EGFR ligands occurs via essential residues within peptidase domains. The study was challenged few months later by (Hosur et al., 2018)on the idea of TACE independent mediated regulation of EGFR ligand. Later in 2018, conditional deletion of ADAM17, in RHBDF2 impaired AREG mediated sebaceous gland enlargement, wound healing and alopecia suggesting ADAM17 is essential for shedding of EGFR ligand (Hosur et al., 2018). Another study on breast cancer has stated iRhoms can regulate proliferation during tumorigenesis via GPCR (G-protein coupled receptor) signaling by transactivation of EGFR signaling (Christova et al., 2013). EGFR transactivation via iRhom1 in Breast Cancer promotes the survival of epithelail tumor (Miyazaki et al., 1998).These pseudoproteases are important for the maturation and trafficking of ADAM17 also known as TACE(TNF-α converting enzyme) to plasma membrane from Endoplasmic reticulum through Golgi apparatus and also linked to the fates of TNF-α and EGFR ligands (Lee et al., 2016).In mammalian cells, growth, proliferation and differentiation is held by EGFR signaling pathways. Important EGFR ligands such as AREG, HB-EGF and EGF binds to EGF receptors(EGFR's) as proproteins and must be cleaved to shed into the extracellular compartment in active form for signaling. But these ligands are membrane tethered and different proteases like active rhomboids and pseudoproteases via client proteins helps in cleaving membrane bound EGFR proligands to convert them into biologically active proteins during proliferation (Dulloo et al., 2019). Moreover, the physiological targets as substrate for enzymes like iRhoms are critical to find but many studies on mice models and D. melanogaster suggest EGF ligands as potential substrates for these pseudo enzymes (Urban and Dickey, 2011).

In order to test the hypothesis of the research, RNA-seq datasets (GSE52194. GSE130660, GSE69240, GSE110114, GSE45419, GSE51124, GSE148991) from GEO were used to find the isoform abundance via new tuxedo pipeline. Furthermore, they were analysed for differential isoform usage using isoformSwitchAnalyzeR. The research reports isoform switching from canonical to the cub transcript at DCIS, grade 2, grade 3 and from cub to canonical at CTCs during neoplastic growth. In order to get insight into its paradoxical role during cell proliferation, its interacting partners were found using correlation and were subjected for enrichment analysis. Neither ERAD nor EGFR signalling pathways were found but the biological processes leading to TACE dependent EGFR pathway were found enriched i.e, MAPK signalling pathway, GPCR pathway and Toll-like receptor pathways. Leading-edge analysis was performed to find the interesting genes across the enriched processes. RELA, RELB, RRAS, GNA12, PRKACA were the gene found among 2/3 of the biological processes for DCIS,grade2 and grade3, while NFKB1 was found in 2/5 of the biological processes for CTCs. Inorder to explain the paradoxical role of iRhom2 as mentioned earlier, we propose a tentative pathway working in the switch hence Cubs pathway as shown in Figure 6. Previously it is known that, TACE is synthesized in endoplasmic reticulum as an immature form containing an inhibitory prodomain that prevents its proteolytic activity. iRhom forms a complex with FRMD8 protein which is an interacting protein . This complex along with an enzyme called furin helps in the removal of the prodomain and converts the TACE into active form in Golgi. The mature TACE trafficks to the plasma membrane with iRhom. The binding of 14-3-3 proteins to iRhom N-terminal domain seems to result in weakening of interaction with TACE at the cell surface. The TACE at the plasma membrane then simultaneously cleaves the EGFR ligands and TNF-α ligands via its sheddase activity thereby mitigating the onset of signaling and inflammatory pathways (Dulloo et al., 2019).Without iRhoms there is no TACE maturation and therefore no TACE activity. Mounting literature and evidence from physiological and molecular data claims the alterations in TACE function due to evident mutations or deletions in N-terminal (Blaydon et al., 2012), (Brooke et al., 2014), (Siggs et al., 2014), (Li et al., 2015).

Evidence suggests that various ADAM metalloproteases (ADAM10, ADAM12, ADAM17) are activated by GPCR agonists to produce mature EGFR ligand leading to EFGR transactivation. EGFR transactivation via GPCR agonist. It has been well documented that EGFR transactivation via GPCR plays a crucial role in proliferation and migration associated physiological functions. GPCR signalling pathway begins with activating the receptors with suitable agonists (ligand activation) and ends with the downstream regulation of various cellular processes such as proliferation, migration, angiogenesis, differentiation, and survival. Activated via agonists (lipids, proteins, amino acids, bio-amines, nucleotides, hormones, or neurotransmitters), GPCRs function by interacting with intracellular G-proteins. G-proteins are guanine nucleotide-binding heterotrimeric proteins, also known as regulatory proteins. They are formed by the combination of three subunits α, β, γ. They are identified via Gα monomers, further

grouped into four families (Gαs, Gαi, Gαq and Gα12. This is the classic GPCR activation mode and further associated with specific effector proteins and require second messengers such as (activation of CA2+, activation of protein kinase C (PKC) and generation of reactive oxygen species (ROS) that amplify downstream pathways (MAPK/ERK/Ras/Raf, P13K/mTOR signalling/Akt).

PRKACA is a protein kinase cAMP activated catalytic subunit alpha and helps in the phosphorylation of different enzymes and proteins. cAMP-dependent phosphorylation of proteins is important to many cellular processes, including differentiation, proliferation, and apoptosis. Several mutations in this gene promote a wide variety of cancers. In addition, PRKACA has been found to regulate resistance to HER2 targeted therapy (Moody et al., 2014). GNA12 is an α guanine nucleotide-binding protein. These heterotrimeric subunits link GPCRs to the nucleotide exchange factors, which in turn interacts with Rho GTPases that regulate cell invasion in breast cancer (Chia et al., 2014). The research also states that the activation of GNA12 in BC stimulates the promotor activity via NF-κB binding of interleukins and matrix metalloproteinase (MMP-2). RELA (v-rel avian reticuloendotheliosis viral oncogene homolog A), also known as p65 transcription factor and NF-κB subunit, is a protein-coding gene. NF-κB is a transcription factor involved in several biological processes like cell growth, inflammation, tumorigenesis, immunity and apoptosis. It is ubiquitous, i.e., present in an inactive form in the cytoplasm by specific inhibitors; upon degradation of these inhibitors, NF-κB moves to the nucleus and regulates specific genes. NF-κB comprises NF-κB1 or NF-κB2 bound to either subunit (Rel like domain-containing proteins) REL, RELA or RELB (Chaturvedi et al., 2011). RELA-NF-κB1 appears to be the most abundant complex. RELA is expressed in many cells like epithelial, neuronal, endothelial, and activation of this gene is positively correlated with multiple cancers. Post-transcriptional modification like methylation is associated with NF-κB1 in Breast cancer (Cancer, 2019). RELB is found to be expressed at higher levels in Breast cancer in regulating the non-canonical NF-κB pathway. It promotes cell proliferation and enhances cell motility by activating epithelial to mesenchymal transition (EMT) (Wang et al., 2020). These signaling pathways are initiated by the binding of extracellular growth factors (ligands/ signaling molecules) to transmembrane receptor tyrosine kinases (RTKs) such as EGFR. RTKs are linked indirectly to Ras via two proteins, GRB2 and Sos.

Ras cycles between an inactive GDP-bound form and active GTP-bound form. Ras cycling requires the assistance of two proteins, GEF and GAP. The SH2 domain in GRB2, an adapter protein, binds to specific phosphotyrosines in activated RTKs. The two SH3 domains in GRB2 then bind Sos, a guaninenucleotide exchange factor, thereby bringing Sos close to membrane-bound Ras· GDP and activating its exchange function. Binding of Sos to inactive Ras causes a large conformational change that permits release of GDP and binding of GTP. RAS becomes active by conversion of GDP to GTP leading to the activation of RAF and MAPK signalling pathway. RRAS is a small GTPase binding protein, and it is involved in angiogenesis, cell adhesion, neuronal regulation and vasculogenesis. Recently, a negative association exists between activation of the RRAS gene and breast cancer progression, and loss of activation of this gene leads to carcinogenesis (Song et al., 2014). The Ras then leads to the activation of RAF and MAPK signalling pathway. MAPK signalling pathway then phosphorylates the iRhom2 N-terminal domain. iRhom2 binding with ADAM17 controls several aspects of its activity, including stimulated shedding activity on the cell surface. ADAM17 shedding stimuli triggers MAP kinase-dependent phosphorylation of iRhom2 N terminal cytoplasmic tail. The regulation of sheddase activity at the cell surface is controlled via several stimulatory agents like G protein-coupled receptors, Toll-like receptors and phorbol esters. (Cavadas et al., 2017) focuses that iRhom2 does not control the trafficking to the cell surface; rather, phosphorylated iRhom2 controls rapid stimulation of TACE activity. (Grieve et al., 2017) showed that GPCR in the presence of histamine agonists triggers TACE dependent release of EGFR ligands like TGF α and amphiregulin in an iRhom2 phosphorylation dependent manner.

Thus, it can be anticipated that during DCIS,grade2 and grade3 the isoform switches from canonical to cub isoform where the cells need to proliferate cancer growth via EGFR signalling pathway and here EGFR pathway is upregulated by all possible means often by indirect activation of TACE by GPCR agonists. Whereas thsi phenomenon starts decreasing when cells undergo metastasises where the primary task is invasion and the isoform switches back to canonical as observed during CTCs aswell.

CONCLUSIONS

Therefore, neither EGFR nor ERAD was found enriched for its interacting partners according to the hypothesis to explain the paradoxical role of RHBDF2. However, pathways leading to TACE depen-

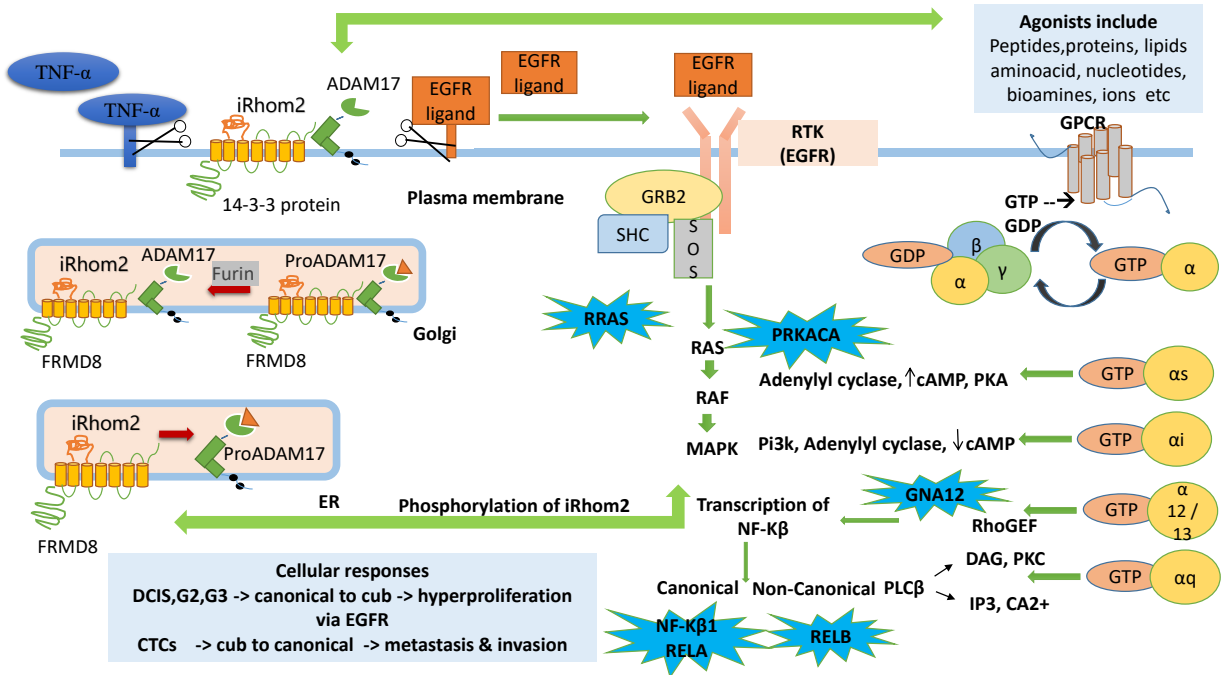

Figure 6. Cubs pathway.The figure shows the proposed mechanistic links between signalling pathways GPCR, MAPK, RTK (EGFR) and the role of ADAM17 (TACE) dependent RHBDF2 (iRhom2) during tumourogenesis. The enriched genes (PRKACA, RRAS, GNA12, NF-κB1, RELA, RELB) in the leading-edge analysis are highlighted in blue bubbles and shown at the site involved during downstream signalling.

dent EGFR signalling pathways were more observant, specifically MAPK signalling pathways, GPCR signalling pathways, and Toll-like receptor pathways in DCIS, grade 2 and grade 3. Nevertheless, it is noteworthy that during CTCs, the cub isoform switchs back to the canonical isoform, so in addition to the processes mentioned above, the Proteasomal degradation pathway and cytoplasmic ribosomal protein pathways were also found significantly enriched. Therefore, it could be inferred that both the isoforms have separate physiological roles to play during tumorigenesis.

ACKNOWLEDGMENTS

The authors would like to thank the Higher Education Commissions Pakistan for funding of the Project. Furthermore would like to thank the research fellows of Data Analytic Labs.

REFERENCES

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bryant, S. H., Canese, K., Church, D. M., Clark, K., DiCuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Geer, L. Y., Gorelenkov, V., Hoeppner, M., Johnson, M., Kelly, C., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Krasnov, S., Kuznetsov, A., Landsman, D., Lipman, D. J., Lu, Z., Madden, T. L., Madej, T., Maglott, D. R., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Ostell, J., O'Sullivan, C., Panchenko, A., Phan, L., Preussm, D., Pruitt, K. D., Rubinstein, W., Sayers, E. W., Schneider, V., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Starchenko, G., Tatusova, T. A., Trawick, B., Vakatov, D., Wang, Y., Ward, M. W., Wilbur, W. J., Yaschenko, E., and Zbicz, K. (2013). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 41:8–20.

Adrain, C. and Freeman, M. (2012). New lives for old: Evolution of pseudoenzyme function illustrated by irhoms. *Nature Reviews Molecular Cell Biology*, 13:489–498.

Al-Salihi, M. A. and Lang, P. A. (2020). Irhom2: An emerging adaptor regulating immunity and disease. *International Journal of Molecular Sciences*, 21:1–20.

Allred, D. C., Mohsin, S. K., and Fuqua, S. A. (2001). Histological and biological evolution of human premalignant breast disease. *Endocrine-Related Cancer*, 8:47–61.

Bergbold, N. and Lemberg, M. K. (2013). Biochimica et biophysica acta emerging role of rhomboid family proteins in mammalian biology and disease ? *BBA - Biomembranes*, 1828:2840–2848.

Blaydon, D. C., Etheridge, S. L., Risk, J. M., Hennies, H. C., Gay, L. J., Carroll, R., Plagnol, V., McRonald, F. E., Stevens, H. P., Spurr, N. K., Bishop, D. T., Ellis, A., Jankowski, J., Field, J. K., Leigh, I. M., South, A. P., and Kelsell, D. P. (2012). Rhbdf2 mutations are associated with tylosis, a familial esophageal cancer syndrome. *American Journal of Human Genetics*, 90:340–346.

Brooke, M. A., Etheridge, S. L., Kaplan, N., Simpson, C., O'Toole, E. A., Ishida-Yamamoto, A., Marches, O., Getsios, S., and Kelsell, D. P. (2014). irhom2-dependent regulation of adam17 in cutaneous disease and epidermal barrier function. *Human Molecular Genetics*, 23:4064–4076.

Cancer, B. (2019). Methylation of the rela gene is associated with. *Molecules (Basel, Switzerland)*, pages 1–9.

Cavadas, M., Oikonomidi, I., Gaspar, C. J., Burbridge, E., Badenes, M., Félix, I., Bolado, A., Hu, T., Bileck, A., Gerner, C., Domingos, P. M., von Kriegsheim, A., and Adrain, C. (2017). Phosphorylation of irhom2 controls stimulated proteolytic shedding by the metalloprotease adam17/tace. *Cell Reports*, 21:745–757.

Chabot, B. and Shkreta, L. (2016). Defective control of pre-messenger rna splicing in human disease. *Journal of Cell Biology*, 212:13–27.

Chaturvedi, M. M., Sung, B., Yadav, V. R., Kannappan, R., and Aggarwal, B. B. (2011). Nf-κb addiction and its role in cancer: One size does not fit all. *Oncogene*, 30:1615–1630.

Chia, C. Y., Kumari, U., and Casey, P. J. (2014). Breast cancer cell invasion mediated by gα12 signaling involves expression of interleukins-6 and -8, and matrix metalloproteinase-2. *Journal of Molecular Signaling*, 9:1–11.

Christova, Y., Adrain, C., Bambrough, P., Ibrahim, A., and Freeman, M. (2013). Mammalian irhoms have distinct physiological functions including an essential role in tace regulation. *EMBO Reports*, 14:884–890.

Di, C., Zhang, Q., Chen, Y., Wang, Y., Zhang, X., Liu, Y., Sun, C., Zhang, H., and Hoheisel, J. D. (2018). Function , clinical application , and strategies of pre-mrna splicing in cancer. *Cell Death and Differentiation*.

Dulloo, I., Muliyil, S., and Freeman, M. (2019). The molecular , cellular and pathophysiological roles of irhom pseudoproteases. *Open Biology*.

Etheridge, S. L., Brooke, M. A., Kelsell, D. P., and Blaydon, D. C. (2013). Rhomboid proteins: A role in keratinocyte proliferation and cancer. *Cell and Tissue Research*, 351:301–307.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016). The pfam protein families database : towards a more sustainable future. *Nucleic acids research*, 44:279–285.

Grieve, A. G., Xu, H., Künzel, U., Bambrough, P., Sieber, B., and Freeman, M. (2017). Phosphorylation of irhom2 at the plasma membrane controls mammalian tace-dependent inflammatory and growth factor signalling. *eLife*, 6:1–22.

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., and Cardoso, F. (2019). *Breast cancer*, volume 5. Nature Reviews Disease Primers.

Hosur, V., Farley, M. L., Burzenski, L. M., Shultz, L. D., and Wiles, M. V. (2018). Adam17 is essential for ectodomain shedding of the egf-receptor ligand amphiregulin. *FEBS Open Bio*, 8:702–710.

Hosur, V., Johnson, K. R., Burzenski, L. M., Stearns, T. M., Maser, R. S., and Shultz, L. D. (2014). Rhbdf2 mutations increase its protein stability and drive egfr hyperactivation through enhanced secretion of amphiregulin. *Proceedings of the National Academy of Sciences of the United States of America*, 111.

jian Kang, Y., chang Yang, D., Kong, L., Hou, M., qi Meng, Y., Wei, L., and Gao, G. (2017). Cpc2 : a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic acids research*, 45:12–16.

Lee, M. Y., Nam, K. H., and Choi, K. C. (2016). irhoms; its functions and essential roles. *Biomolecules*

and Therapeutics, 24:109–114.

Li, X., Maretzky, T., Weskamp, G., Monette, S., Qing, X., Issuree, P. D. A., Crawford, H. C., McIlwain, D. R., Mak, T. W., Salmon, J. E., and Blobel, C. P. (2015). irhoms 1 and 2 are essential upstream regulators of adam17-dependent egfr signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 112:6080–6085.

Lyu, Z., Mao, Z., Li, Q., Xia, Y., Liu, Y., He, Q., Wang, Y., Zhao, H., Lu, Z., and Zhou, Q. (2018). Pparγ maintains the metabolic heterogeneity and homeostasis of renal tubules. *EBioMedicine*, 38:178–190.

Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., and Baradaran, B. (2017). The different mechanisms of cancer drug resistance : A brief review. *Tabriz University of Medical Sciences*, 7:339–348.

Maza, E. (2016). In papyro comparison of tmm (edger), rle (deseq2), and mrn normalization methods for a simple two-conditions-without-replicates rna-seq experimental design. *Frontiers in Genetics*, 7:1–8.

Miyazaki, M., Lamharzi, N., Schally, A. V., Halmos, G., Szepeshazi, K., Groot, K., and Cai, R. Z. (1998). Original paper inhibition of growth of mda-mb-231 human breast cancer xenografts in nude mice by bombesin / gastrin-releasing peptide (grp) antagonists rc-3940-ii and rc-3095. *European journal of cancer (Oxford, England:1990)*, 34:710–717.

Moody, S. E., Schinzel, A. C., Singh, S., Izzo, F., Strickland, M. R., Luo, L., Thomas, S. R., Boehm, J. S., Kim, S. Y., Wang, Z. C., and Hahn, W. C. (2014). Prkaca mediates resistance to her2-targeted therapy in breast cancer cells and restores anti-apoptotic signaling. *Oncogene*, 34:2061–2071.

Nicholson, R. I., Gee, J. M. W., and Harper, M. E. (2001). Egfr and cancer prognosis. *European journal of cancer (Oxford, England:1990)*, 37:9–15.

Nickel, K. A., Halper, J., and Moses, H. L. (1983). Transforming growth factors in solid human malignant neoplasms1. *Cancer research*, pages 1966–1971.

Park, K. C., Dharmasivam, M., and Richardson, D. R. (2020). The role of extracellular proteases in tumor progression and the development of innovative metal ion chelators that inhibit their activity. *International journal of molecular sciences*.

Pertea, G. and Pertea, M. (2020). Gff utilities: Gffread and gffcompare. *F1000Research*, 9:1–19.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of rna-seq experiments with hisat, stringtie and ballgown. *Nature Protocols*, 11:1650–1667.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). Seqkit : A cross-platform and ultrafast toolkit for fasta / q file manipulation. *PLOS ONE*, pages 1–10.

Siggs, O. M., Grieve, A., Xu, H., Bambrough, P., Christova, Y., and Freeman, M. (2014). Genetic interaction implicates irhom2 in the regulation of egf receptor signalling in mice. *Biology Open*, 3:1151–1157.

Soneson, C., Love, M. I., and Robinson, M. D. (2020). Differential analyses for rna-seq : transcript-level estimates improve gene-level inferences [version 2 ; peer review : 2 approved]. *F1000Research*, pages 1–23.

Song, J., Zheng, B., Bu, X., Fei, Y., and Shi, S. (2014). Negative association of r-ras activation and breast cancer development. *Oncology Reports*, 31:2776–2784.

Sorek, R., Shamir, R., and Ast, G. (2004). How prevalent is functional alternative splicing in the human genome? *Trends in Genetics*, 20:68–71.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102:15545–15550.

Urban, S. and Dickey, S. W. (2011). The rhomboid protease family : a decade of progress on function and mechanism. *Genome Biology*.

Wang, M., Zhang, Y., Xu, Z., Qian, P., Sun, W., Wang, X., Jian, Z., Xia, T., Xu, Y., and Tang, J. (2020). Relb sustains endocrine resistant malignancy: An insight of noncanonical nf-κb pathway into breast cancer progression. *Cell Communication and Signaling*, 18:1–17.

xiang Lu, Z., Huang, Q., Park, J. W., Shen, S., Lin, L., Tokheim, C. J., Henry, M. D., and Xing, Y. (2015). Transcriptome-wide landscape of pre-mrna alternative splicing associated with metastatic colonization. *Molecular cancer research: MCR*, pages 305–319.

Zettl, M., Adrain, C., Strisovsky, K., Lastun, V., and Freeman, M. (2011). Rhomboid family pseudoproteases use the er quality control machinery to regulate intercellular signaling. *Cell*, 145:79–91.

547 Zhang, Z. F., Pal, S., Bi, Y., Tchou, J., and Davuluri, R. V. (2013). Isoform level expression profiles
548 provide better cancer signatures than gene level expression profiles. *Genome Medicine*, 5.