Chapter 3

3. Objective I.: Study the expression pattern of pro-apoptotic, anti-apoptotic and proliferation pathway genes on treatment with recombinant human TRAIL (rhTRAIL)

3.1. Background of the study

Breast cancer is the most potent nemesis of women's health worldwide, every year as per WHO 2.1 million new cases are registered[150,151]. Till date chemotherapy is the main stream method applied to treat breast cancer patients, even though other therapies are considered for the treatment of the patients such as targeted hormonal receptor therapy but the effectiveness of these methods are not as much as they were hoped for treating the growing cancer[58]. In search of a more efficient way for treatment of breast cancer, TRAIL (Tumor necrosis factor-related apoptosis-inducing ligand) is studied extensively in the last two decades[55,152]. TRAIL is well known for its unique ability to induce apoptosis in a tissue specific manner, this specific ability of TRAIL makes it a potent anti-cancer agent and it has been proven in pre-clinical trials that TRAIL is a secure and effective way of treating breast cancer patients[55,56]. Certain short-comes have proven clinical trial of TRAIL less optimal than expected, the short-life of TRAIL and certain tumors showing resistance towards TRAIL were the major road block against the expected outcome of TRAIL[113,153,154]. Therefore, identifying the different biomarkers of apoptosis inducing pathway of TRAIL is a primary requirement for enhancing TRAIL efficiency as an anti-cancer agent.

Biomarkers can be defined as those biological substances which can be used to predict medical state. Biomarkers are nowadays used extensively in drug development and for understanding the relationship between biological pathways and clinical outcomes[155]. TRAIL has been identified as a biomarker for cardiovascular diseases for its beneficial therapeutic role in treatment of diabetes and cardiovascular diseases[57].

TRAIL mediated modulation of our immune system has been one of the key features of TRAIL signalling. One of the main component of innate and adaptive immune system is the 'Complement system', which can activated by classical pathway (antigen-antibody complex formation system), lectin pathway and/or alternative pathway[53,54]. Initiation of the classical pathway is regulated by C1q, C1r, C1s, C2, C4, CFH; lectin pathway starts with MBL, MASP2, C2, C4; and the alternative pathway involves C3, CFB, CFP, CFH, CFI, CD59. All the three pathways convert inactive C5 into the active C5 convertase, which finally generates the MAC (membrane attack complex) resulting in cell lysis[54]. Several studies have shown the complement system to be activated in response to tumor antigens[52,54]In our study we do find an interesting link between complement system and TRAIL, which has not been reported before.

Advancement in the field of bioinformatic analysis has provide us with high throughput genomic data including genomics, proteomics and metabolomics data. This has enabled us a significant insight in identifying pathways effected in different medical conditions and to understand molecular mechanisms responsible for various diseases.

3.2. Results

3.2.1 Identification of genes that were differentially expressed at the transcript level between the TRAIL treated and control MDA-MB-231 breast cancer cells

Breast cancer cell line MDA-MB-231 was treated with 50ng/ml rhTRAIL and parallel mock treatment was given as control. In performed trypan blue assay we observed increased cell death due to rhTRAIL treatment (Fig. 1a) and even cleaved caspase-3 level was elevated in comparison to the control MDA-MB-231 cells (Fig. 1b). Microarray was performed to identify the genes involved in TRAIL induced apoptosis in the treated cells. Microarray analysis showed a total of 32,045 probes, out of which 4721 probes were found to be significantly different between the rhTRAIL and mock treated control cells. 3633 probes out of 4721 probes consisted of transcripts having an NM ID (Genbank accession number), among which we only observed 3463 number of unique transcripts (transcripts have unique NM ID) (Fig. 1c). Next, shortlisting of the transcripts were done with a cut off of fold change of greater than 1.5 and less than -1.5. Finally, we identified 303

transcripts that were differentially expressed between the rhTRAIL treated and control MDA-MB-231 breast cancer cells. 297 out of 303 transcripts were unique differentially expressed genes (DEGs) (One of the copies of the 6 duplicate DEGs were removed)., 200 DEGs were observed to be upregulated and 97 DEGs were found to be downregulated in the rhTRAIL treated cells in comparison to the control cells, out of these 297 DEGs (Fig. 1d, Table 3.1, 3.2).

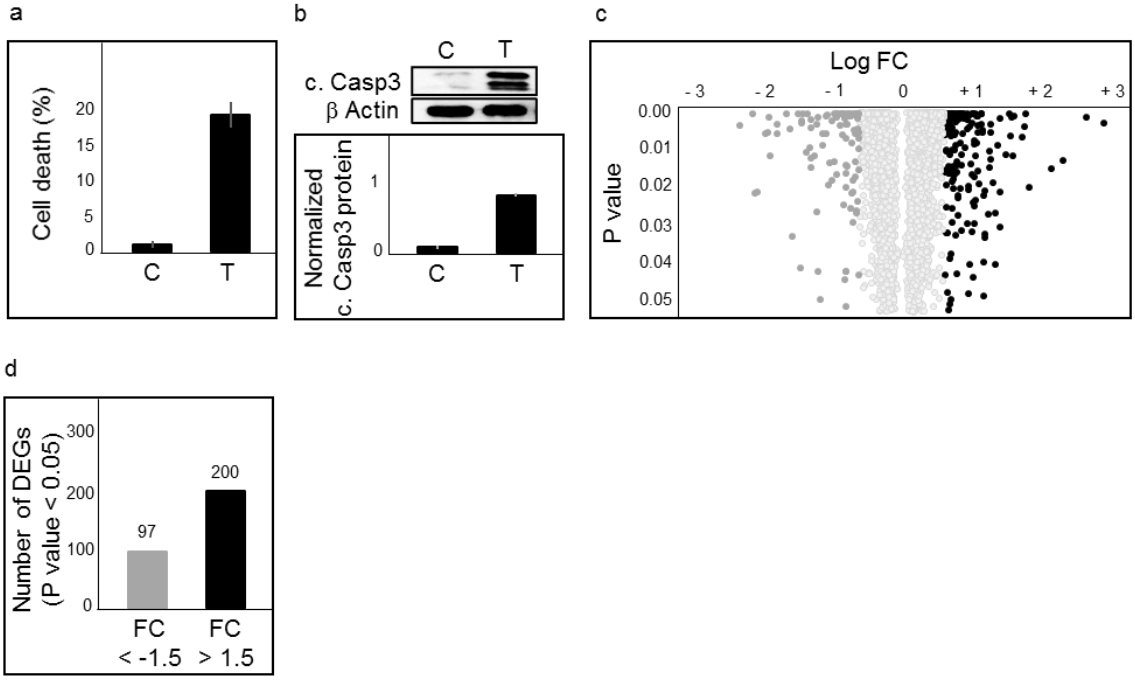

Figure 3.1. *rhTRAIL induced cytotoxicity and identification DEGs from microarray data. (a) Cell death was measured using trypan blue assay in MDA-MB-231 with 50ng/ml rhTRAIL treatment. (b) Cleaved caspase-3 protein was measured using western blotting. β-actin was used as an endogenous control. (c) Number of upregulated and downregulated DEGs with a fold change of >1.5, < -1.5. (d) Fold change of 297 selected DEGs (C: control; Tr: rhTRAIL; * indicates p≤0.05).*

Table 3.1: Genes that are upregulated in rhTRAIL treated cells compared to control cells

Gene Symbol	Genbank Accession	Gene Name
TNXB	NM_032470	tenascin XB
ATP2A3	NM_174953	ATPase, Ca++ transporting, ubiquitous
UNC5B	NM_170744	unc-5 homolog B (C. elegans)
ADAMTS13	NM_139025	ADAM metallopeptidase with thrombospondin type 1 motif, 13

Gene Symbol	Genbank Accession	Gene Name
FN1	NM_054034	fibronectin 1
PCM1	NM_006197	pericentriolar material 1
PLXNB3	NM_00116325	plexin B3
FBLN7	NM_153214	fibulin 7
SCN1B	NM_199037	sodium channel, voltage gated, type I beta subunit
C1S	NM_201442	complement component 1, s subcomponent
TNFRSF11	NM_002546	tumor necrosis factor receptor superfamily, member
CPE	NM_001873	carboxypeptidase E
HSPA1A	NM_005345	heat shock 70kDa protein 1A
CFH	NM_000186	complement factor H
BMP8A	NM_181809	bone morphogenetic protein 8a
CFP	NM_002621	complement factor properdin
CD69	NM_001781	CD69 molecule
TMC8	NM_152468	transmembrane channel-like 8
H6PD	NM_004285	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)
ARHGAP20	NM_020809	Rho GTPase activating protein 20
ZDHHC8	NM_013373	zinc finger, DHHC-type containing 8
AP5B1	NM_138368	adaptor-related protein complex 5, beta 1 subunit
SERPINB3	NM_006919	serpin peptidase inhibitor, clade B (ovalbumin),
SLC13A3	NM_00101155	solute carrier family 13 (sodium-dependent
EFNB3	NM_001406	ephrin-B3
FBXO2	NM_012168	F-box protein 2
TMEM63C	NM_020431	transmembrane protein 63C
SHC2	NM_012435	SHC (Src homology 2 domain containing)
CIDEB	NM_014430	cell death-inducing DFFA-like effector b
VIPR1	NM_004624	vasoactive intestinal peptide receptor 1
CHI3L1	NM_001276	chitinase 3-like 1 (cartilage glycoprotein-39)
PLEKHB1	NM_021200	pleckstrin homology domain containing, family B
PDE4DIP	NM_00119883	phosphodiesterase 4D interacting protein
ETV3	NM_00114531	ets variant 3
METTL9	NM_016025	methyltransferase like 9
SPNS3	NM_182538	spinster homolog 3 (Drosophila)
PCDHGA5	NM_032054	protocadherin gamma subfamily A, 5
FOLR1	NM_016725	folate receptor 1 (adult)
MPPE1	NM_023075	metallophosphoesterase 1
LEFTY1	NM_020997	left-right determination factor 1
PCDHB9	NM_019119	protocadherin beta 9
TG	NM_003235	thyroglobulin
PTGDS	NM_000954	prostaglandin D2 synthase 21kDa (brain)
TAS2R30	NM_00109764	taste receptor, type 2, member 30
TIGD6	NM_030953	tigger transposable element derived 6
C12orf54	NM_152319	chromosome 12 open reading frame 54
BCKDHA	NM_000709	branched chain keto acid dehydrogenase E1, alpha
TPP1	NM_000391	tripeptidyl peptidase I
ZDHHC1	NM_013304	zinc finger, DHHC-type containing 1

Gene Symbol	Genbank Accession	Gene Name
ADSSL1	NM_199165	adenylosuccinate synthase like 1
SPRY1	NM_00125803	sprouty homolog 1, antagonist of FGF signaling
AGT	NM_000029	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)
CALML6	NM_138705	calmodulin-like 6
CAMP	NM_004345	cathelicidin antimicrobial peptide
CA12	NM_001218	carbonic anhydrase XII
THSD4	NM_024817	thrombospondin, type I, domain containing 4
SLA	NM_00104555	Src-like-adaptor
MAN2A2	NM_006122	mannosidase, alpha, class 2A, member 2
SLC1A1	NM_004170	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member
CRIM1	NM_016441	cysteine rich transmembrane BMP regulator 1
C4B	NM_00100202	complement component 4B (Chido blood group)
PAG1	NM_018440	phosphoprotein membrane anchor with
EXD3	NM_00128682	exonuclease 3'-5' domain containing 3
SEC16B	NM_033127	SEC16 homolog B (S. cerevisiae)
PMEPA1	NM_020182	prostate transmembrane protein, androgen induced 1
MUC1	NM_002456	mucin 1, cell surface associated
CFAP70	NM_145170	cilia and flagella associated protein 70
HPR	NM_020995	haptoglobin-related protein
VPREB3	NM_013378	pre-B lymphocyte 3
CAMK2B	NM_172082	calcium/calmodulin-dependent protein kinase II beta
PACS2	NM_00110091	phosphofurin acidic cluster sorting protein 2
GOLGA6C	NM_00116440	golgin A6 family, member C
TMEM120B	NM_00108082	transmembrane protein 120B
CDH1	NM_004360	cadherin 1, type 1, E-cadherin (epithelial)
NF1	NM_000267	neurofibromin 1
AK8	NM_152572	adenylate kinase 8
CA9	NM_001216	carbonic anhydrase IX
CFI	NM_000204	complement factor I
CD14	NM_00117410	CD14 molecule
NFAM1	NM_145912	NFAT activating protein with ITAM motif 1
TRAPPC6A	NM_024108	trafficking protein particle complex 6A
IL2RB	NM_000878	interleukin 2 receptor, beta
THBS2	NM_003247	thrombospondin 2
SEC14L6	NM_00119333	SEC14-like 6 (S. cerevisiae)
SYT12	NM_177963	synaptotagmin XII
ZP1	NM_207341	zona pellucida glycoprotein 1 (sperm receptor)
CMKLR1	NM_00114234	chemokine-like receptor 1
GGT5	NM_00109978	gamma-glutamyltransferase 5
KLHL42	NM_020782	kelch-like family member 42
ODF4	NM_153007	outer dense fiber of sperm tails 4
ITGB3	NM_000212	integrin, beta 3 (platelet glycoprotein IIIa, antigen
RBM14	NM_006328	RNA binding motif protein 14

Gene Symbol	Genbank Accession	Gene Name
HMGA2	NM_003483	high mobility group AT-hook 2
RGS9	NM_003835	regulator of G-protein signaling 9
PER2	NM_022817	period circadian clock 2
MLXIPL	NM_032951	MLX interacting protein-like
C2	NM_00128245	complement component 2
GRIK2	NM_00116624	glutamate receptor, ionotropic, kainate 2
HP	NM_005143	haptoglobin
NOXA1	NM_006647	NADPH oxidase activator 1
RBP1	NM_002899	retinol binding protein 1, cellular
GAB3	NM_00108157	GRB2-associated binding protein 3
NUCB1	NM_006184	nucleobindin 1
NR3C1	NM_00120426	nuclear receptor subfamily 3, group C, member 1
ABCC3	NM_003786	ATP-binding cassette, sub-family C (CFTR/MRP),
PRICKLE2	NM_198859	prickle homolog 2 (Drosophila)
OTX1	NM_014562	orthodenticle homeobox 1
SLC38A3	NM_006841	solute carrier family 38, member 3
CD59	NM_203330	CD59 molecule, complement regulatory protein
UBD	NM_006398	ubiquitin D
STEAP2	NM_00124494	STEAP family member 2, metalloreductase
FAM177B	NM_207468	family with sequence similarity 177, member B
CFB	NM_001710	complement factor B
VAPB	NM_004738	VAMP (vesicle-associated membrane protein)-associated protein B and C
PSAPL1	NM_00108538	prosaposin-like 1 (gene/pseudogene)
COL5A1	NM_000093	collagen, type V, alpha 1
PHGDH	NM_006623	phosphoglycerate dehydrogenase
LYPD1	NM_144586	LY6/PLAUR domain containing 1
KIAA0368	NM_00108039	KIAA0368
PLA2G4C	NM_003706	phospholipase A2, group IVC (cytosolic, calcium-independent)
CCDC85A	NM_00108043	coiled-coil domain containing 85A
GPC6	NM_005708	glypican 6
CPD	NM_001304	carboxypeptidase D
ITIH4	NM_002218	inter-alpha-trypsin inhibitor heavy chain family, member 4
DLX1	NM_178120	distal-less homeobox 1
STRA6	NM_00119904	stimulated by retinoic acid 6
RARRES1	NM_206963	retinoic acid receptor responder (tazarotene induced)
MEIS3	NM_00100981	Meis homeobox 3
SERPINA5	NM_000624	serpin peptidase inhibitor, clade A (alpha-1
CTSD	NM_001909	cathepsin D
UTRN	NM_007124	utrophin
NECAB2	NM_019065	N-terminal EF-hand calcium binding protein 2
NTNG2	NM_032536	netrin G2
SORCS2	NM_020777	sortilin-related VPS10 domain containing receptor 2

Gene Symbol	Genbank Accession	Gene Name
APOE	NM_00130268	apolipoprotein E
C1RL	NM_016546	complement component 1, r subcomponent-like
ZNF365	NM_014951	zinc finger protein 365
MORN3	NM_173855	MORN repeat containing 3
BBS1	NM_024649	Bardet-Biedl syndrome 1
MMP11	NM_005940	matrix metallopeptidase 11 (stromelysin 3)
PLEKHG1	NM_00102988	pleckstrin homology domain containing, family G
FBXL22	NM_203373	F-box and leucine-rich repeat protein 22
AZGP1	NM_001185	alpha-2-glycoprotein 1, zinc-binding
PIGZ	NM_025163	phosphatidylinositol glycan anchor biosynthesis,
SSPN	NM_005086	sarcospan
ABCA3	NM_001089	ATP-binding cassette, sub-family A (ABC1),
KLRC3	NM_007333	killer cell lectin-like receptor subfamily C, member 3
MAP2	NM_002374	microtubule-associated protein 2
FECH	NM_00101251	ferrochelatase
CA11	NM_001217	carbonic anhydrase XI
OLA1	NM_013341	Obg-like ATPase 1
ASS1	NM_000050	argininosuccinate synthase 1
CEMIP	NM_018689	cell migration inducing protein, hyaluronan binding
P2RY12	NM_022788	purinergic receptor P2Y, G-protein coupled, 12
C9orf9	NM_018956	chromosome 9 open reading frame 9
ABI3BP	NM_015429	ABI family, member 3 (NESH) binding protein
NFIB	NM_005596	nuclear factor I/B
CDHR3	NM_152750	cadherin-related family member 3
COX18	NM_00129773	COX18 cytochrome c oxidase assembly factor
FAM110C	NM_00107771	family with sequence similarity 110, member C
CHI3L2	NM_00102519	chitinase 3-like 2
LTBP3	NM_021070	latent transforming growth factor beta binding
ARHGAP4	NM_00116474	Rho GTPase activating protein 4
KRT86	NM_002284	keratin 86, type II
SPINT1	NM_181642	serine peptidase inhibitor, Kunitz type 1
ZMAT3	NM_022470	zinc finger, matrin-type 3
FUNDC2	NM_023934	FUN14 domain containing 2
LTBP2	NM_000428	latent transforming growth factor beta binding
DSEL	NM_032160	dermatan sulfate epimerase-like
KCNQ1	NM_000218	potassium channel, voltage gated KQT-like
SNED1	NM_00108043	sushi, nidogen and EGF-like domains 1
COL6A2	NM_058174	collagen, type VI, alpha 2
HLA-DQA1	NM_002122	major histocompatibility complex, class II, DQ alpha
SMPDL3B	NM_00100956	sphingomyelin phosphodiesterase, acid-like 3B
CLSTN3	NM_014718	calsyntenin 3
GPR124	NM_032777	G protein-coupled receptor 124
SEL1L3	NM_015187	sel-1 suppressor of lin-12-like 3 (C. elegans)
TCEA2	NM_003195	transcription elongation factor A (SII), 2
MBOAT2	NM_138799	membrane bound O-acyltransferase domain
IL17A	NM_002190	interleukin 17A

Gene Symbol	Genbank Accession	Gene Name
MMP7	NM_002423	matrix metallopeptidase 7 (matrilysin, uterine)
DUOX1	NM_017434	dual oxidase 1
ABCA2	NM_001606	ATP-binding cassette, sub-family A (ABC1),
ASNS	NM_001673	asparagine synthetase (glutamine-hydrolyzing)
COLEC12	NM_130386	collectin sub-family member 12
FOXS1	NM_004118	forkhead box S1
NALCN	NM_052867	sodium leak channel, non selective
PCDHB10	NM_018930	protocadherin beta 10
GRIK4	NM_014619	glutamate receptor, ionotropic, kainate 4
VSTM2L	NM_080607	V-set and transmembrane domain containing 2 like
SERPINA3	NM_001085	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3
RPL27A	NM_000990	ribosomal protein L27a
CTSO	NM_001334	cathepsin O
SNX2	NM_00127819	sorting nexin 2
C1R	NM_001733	complement component 1, r subcomponent
HLA-DQB1	NM_001243961	major histocompatibility complex, class II, DQ beta 1
CXCL10	NM_001565	chemokine (C-X-C motif) ligand 10
UACA	NM_001008224	uveal autoantigen with coiled-coil domains and ankyrin repeats
NOTCH2N	NM_203458	notch 2 N-terminal like
ANTXRL	NM_00127868	anthrax toxin receptor-like

Table 3.2: Genes that are downregulated in rhTRAIL treated cells compared to control cells

Gene	Genbank	Gene Name
WNK2	NM_006648	WNK lysine deficient protein kinase 2
SCUBE3	NM_152753	signal peptide, CUB domain, EGF-like 3
AIM1L	NM_001039	absent in melanoma 1-like
GDF15	NM_004864	growth differentiation factor 15
IQUB	NM_178827	IQ motif and ubiquitin domain containing
JDP2	NM_130469	Jun dimerization protein 2
IFNE	NM_176891	interferon, epsilon
VSTM1	NM_198481	V-set and transmembrane domain containing 1
STMN2	NM_007029	stathmin 2
SP140	NM_001005	SP140 nuclear body protein
WDR87	NM_001291	WD repeat domain 87
TCF7	NM_003202	transcription factor 7 (T-cell specific, HMG-box)
DDIT4	NM_019058	DNA-damage-inducible transcript 4
SLC26A5	NM_206883	solute carrier family 26 (anion exchanger), member 5
CCL3	NM_002983	chemokine (C-C motif) ligand 3

Gene	Genbank	Gene Name
ID4	NM_001546	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
MMP1	NM_002421	matrix metallopeptidase 1 (interstitial collagenase)
SESN2	NM_031459	sestrin 2
DGKK	NM_001013	diacylglycerol kinase, kappa
GLRX	NM_002064	glutaredoxin (thioltransferase)
ATF3	NM_001040	activating transcription factor 3
RNF121	NM_018320	ring finger protein 121
MS4A7	NM_021201	membrane-spanning 4-domains, subfamily A, member 7
PKD1L2	NM_001076	polycystic kidney disease 1-like 2 (gene/pseudogene)
KRTAP2-	NM_001165	keratin associated protein 2-3
USF1	NM_007122	upstream transcription factor 1
GAGE7	NM_021123	G antigen 7
C1QTNF9	NM_001014	C1QTNF9B antisense RNA 1
FGF1	NM_000800	fibroblast growth factor 1 (acidic)
DDIT3	NM_004083	DNA-damage-inducible transcript 3
PDE7B	NM_018945	phosphodiesterase 7B
VAV1	NM_005428	vav 1 guanine nucleotide exchange factor
IL24	NM_001185	interleukin 24
ELF3	NM_004433	E74-like factor 3 (ets domain transcription factor, epithelial-specific)
M1AP	NM_138804	meiosis 1 associated protein
LY6K	NM_017527	lymphocyte antigen 6 complex, locus K
IL1A	NM_000575	interleukin 1, alpha
JAKMIP3	NM_001105	Janus kinase and microtubule interacting protein 3
LOC7284	NM_001294	uncharacterized LOC728485
TRIM29	NM_012101	tripartite motif containing 29
HMMR	NM_012484	hyaluronan-mediated motility receptor (RHAMM)
CGN	NM_020770	cingulin
TRIM61	NM_001012	tripartite motif containing 61
LMO7DN	NM_001257	LMO7 downstream neighbor
NIPAL3	NM_020448	NIPA-like domain containing 3
CTH	NM_001902	cystathionine gamma-lyase
ADM2	NM_024866	adrenomedullin 2
C1orf111	NM_182581	chromosome 1 open reading frame 111
HIST2H3	NM_001005	histone cluster 2, H3a
CT62	NM_001102	cancer/testis antigen 62
ANXA10	NM_007193	annexin A10
ALDH4A	NM_170726	aldehyde dehydrogenase 4 family, member A1
AOAH	NM_001637	acyloxyacyl hydrolase (neutrophil)
C16orf93	NM_001195	chromosome 16 open reading frame 93
C5orf34	NM_198566	chromosome 5 open reading frame 34
RCSD1	NM_052862	RCSD domain containing 1
PRSS45	NM_199183	protease, serine, 45
RAB33A	NM_004794	RAB33A, member RAS oncogene family

Gene	Genbank	Gene Name
CCDC103	NM_001258	coiled-coil domain containing 103
S1PR4	NM_003775	sphingosine-1-phosphate receptor 4
PTPN7	NM_080588	protein tyrosine phosphatase, non-receptor type 7
SLC30A2	NM_001004	solute carrier family 30 (zinc transporter), member 2
PALD1	NM_014431	phosphatase domain containing, paladin 1
KRT34	NM_021013	keratin 34, type I
TEX40	NM_001039	testis expressed 40
EXOC3L4	NM_001077	exocyst complex component 3-like 4
FAM49A	NM_030797	family with sequence similarity 49, member A
GRAMD1	NM_001286	GRAM domain containing 1B
CHAC1	NM_024111	ChaC glutathione-specific gamma-glutamylcyclotransferase 1
MPP4	NM_033066	membrane protein, palmitoylated 4 (MAGUK p55 subfamily member 4)
KIF12	NM_138424	kinesin family member 12
UBE2E1	NM_003341	ubiquitin-conjugating enzyme E2E 1
PVRL1	NM_203286	poliovirus receptor-related 1 (herpesvirus entry mediator C)
NPPB	NM_002521	natriuretic peptide B
LRRC48	NM_031294	leucine rich repeat containing 48
DMBT1	NM_007329	deleted in malignant brain tumors 1
PIK3R5	NM_014308	phosphoinositide-3-kinase, regulatory subunit 5
GORAB	NM_001146	golgin, RAB6-interacting
HIST1H2	NM_021064	histone cluster 1, H2ag
LRP1	NM_002332	low density lipoprotein receptor-related protein 1
PCDH11X	NM_032968	protocadherin 11 X-linked
CCL3L3	NM_001001	chemokine (C-C motif) ligand 3-like 3
ST18	NM_014682	suppression of tumorigenicity 18, zinc finger
TRIB3	NM_021158	tribbles pseudokinase 3
ERICH6	NM_152394	glutamate-rich 6
NRP2	NM_201266	neuropilin 2
KCNRG	NM_199464	potassium channel regulator
NOX1	NM_013955	NADPH oxidase 1
ZNF165	NM_003447	zinc finger protein 165
C16orf97	NM_001242	chromosome 16 open reading frame 97
GSG1	NM_001080	germ cell associated 1
ARTN	NM_057090	artemin
SLC51B	NM_178859	solute carrier family 51, beta subunit
HERPUD	NM_014685	homocysteine-inducible, endoplasmic reticulum stress-
IL11	NM_000641	interleukin 11
TROAP	NM_001100	trophinin associated protein
HRK	NM_003806	harakiri, BCL2 interacting protein

3.2.2. Gene Ontology and KEGG pathway analysis of the 297 differentially expressed transcripts through DAVID gene functional classification tool

For understanding of the basic biological importance of the 297 candidate transcripts, DAVID online algorithm was used to perform GO analysis of these respective genes. We observed that the significant Biological Processes (BP), GO:0030449~regulation of complement activation, has a fold enrichment of 15.1. This GO term comprises of seven genes, namely, CFH (NM_000186, NM_001014975) CFP (NM_002621), C4B (NM_001002029), CFI (NM_000204), C2 (NM_000063, NM_001282458), CD59 (NM_203330) and CFB (NM_001710). GO terms include 'blood microparticle', 'proteinaceous extracellular matrix', 'extracellular matrix', 'extracellular space', 'extracellular region' and 'extracellular exosome' were found to be significant among the Cellular Components (CC). GO terms like 'heparin binding', 'serine-type endopeptidase activity' and 'calcium ion binding' were found significant for Molecular Function (MF) (Fig. 2a, Table. 3.3). We also found that the genes included in the BP GO term 'regulation of complement activation' were also a part of the significant CC and MF as shown in (Fig. 2b). KEGG pathway enrichment analysis result showed hsa04610:Complement and coagulation cascades and hsa05150:Staphylococcus aureus infection have a significant enrichment values of 7- and 9-fold, respectively, in treated rhTRAIL cells. Few complement genes like C1R, C1S, C2, C4B, CFB, CFH, CFI were found to be common between the significant KEGG pathways (Fig. 2c, Table. 3.4).

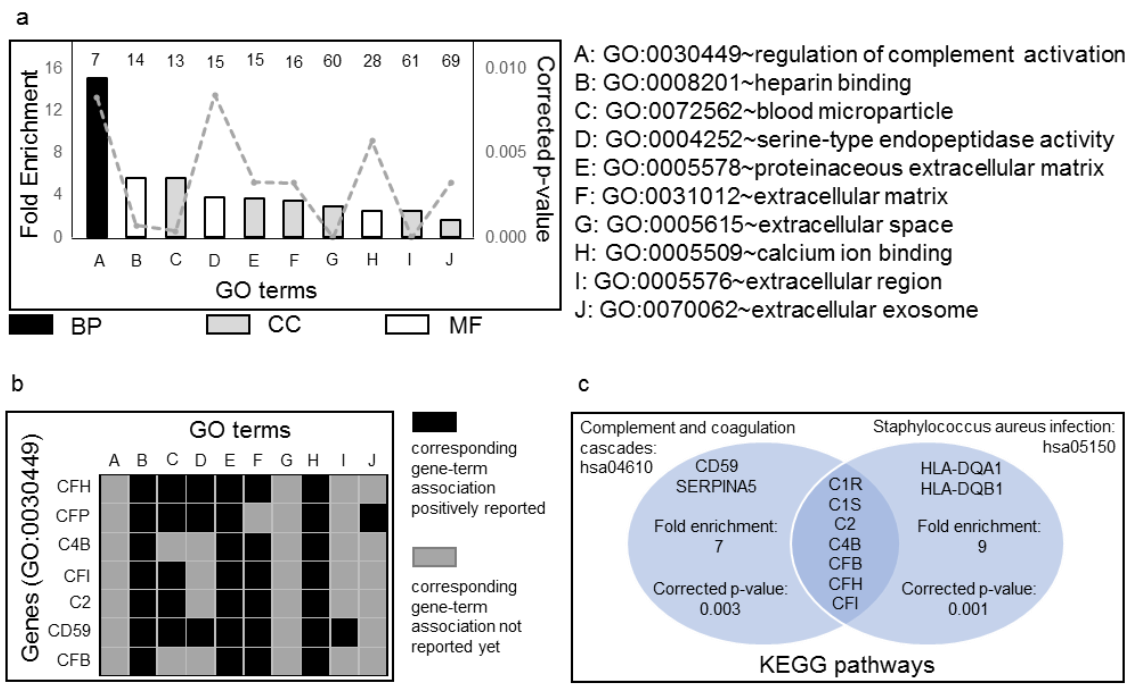

Figure 3.2. *GO enrichment analysis of 297 DEGs and Venn diagram of KEGG pathway analysis. (a) Bar diagram of Significant GO terms (b) Heat map of the candidate DEGs present in the most enriched GO term in the other significant GO terms. (c) Venn diagram of common DEGs between two significant KEGG pathways*

Table 3.3: Top 10 significant GO terms based on 297 differentially expressed genes.

Sr. No	GO terms	GO category	Gene Count	Fold Enrichment	Gene Name
1	GO:0030449~regulation of complement activation	BP	7	15.1	C4B, CFH, CFI, CD59, CFP, CFB, C2
2	GO:0008201~heparin binding	MF	14	5.6	FBLN7, NRP2, TNXB, MMP7, CFH, FN1, LTBP2, FGF1, THBS2, SERPINA5, CXCL10, COL5A1, ABI3BP, APOE
3	GO:0072562~blood microparticle	CC	13	5.6	SERPINA3, ITIH4, CFH, C1S, C1R,

Sr. No	GO terms	GO category	Gene Count	Fold Enrichment	Gene Name
					HP, FN1, HPR, AGT, C4B, APOE, CFB, HSPA1A
4	GO:0004252~serine-type endopeptidase activity	MF	15	3.7	MMP7, C1S, C1R, MMP1, CFI, HP, HPR, PRSS45, C2, C4B, MMP11, C1RL, TPP1, CTSD, CFB
5	GO:0005578~proteinaceous extracellular matrix	CC	15	3.6	FBLN7, TNXB, MMP7, MMP1, FN1, LTBP2, TNFRSF11B, FGF1, MMP11, ADAMTS13, COL5A1, COL6A2, CHI3L1, ZP1, GPC6
6	GO:0031012~extracellular matrix	CC	16	3.5	TNXB, MMP7, MMP1, FN1, LTBP2, LTBP3, CFP, THBS2, THSD4, MMP11, COL5A1, ABI3BP, COL6A2, ZP1, APOE, CTSD
7	GO:0005615~extracellular space	CC	60	2.9	SERPINA3, IL24, HP, FGF1, SERPINA5, NPPB, C4B, VPREB3, CTSO, DMBT1, C1RL, VSTM1, PTGDS, CTSD, CAMP, SERPINB3, IL11, LEFTY1, MMP7, KRT34, TMC8, IL1A, AZGP1, SPINT1, COL6A2, IFNE, CHI3L2, CHI3L1, SMPDL3B, CFB, TNXB, CFH, CCL3L3, CFI, LTBP2, TNFRSF11B, KRT86, CFP, C2, MUC1, ADAMTS13,

Sr. No	GO terms	GO category	Gene Count	Fold Enrichment	Gene Name
					ABI3BP, CCL3, CD59, CD14, APOE, GPC6, PSAPL1, GDF15, ABCA3, FN1, BMP8A, AGT, CXCL10, TG, ARTN, CPD, NUCB1, CPE, IL17A
8	GO:0005509~calcium ion binding	MF	28	2.5	FBLN7, NECAB2, SNED1, LRP1, C1S, C1R, CLSTN3, CALML6, PCDH11X, NOTCH2NL, LTBP2, PKD1L2, LTBP3, THBS2, SCUBE3, DUOX1, ADAMTS13, CDH1, PCDHGA5, MMP1, ANXA10, PCDHB10, MMP11, SYT12, CDHR3, NUCB1, CD69, PCDHB9
9	GO:0005576~extracellular region	CC	61	2.5	SERPINA3, LY6K, NRP2, IL24, HP, FGF1, SERPINA5, NPPB, C4B, DMBT1, CDH1, PTGDS, CTSD, SCN1B, CAMP, IL11, MMP7, MMP1, VSTM2L, MMP11, IL1A, AZGP1, HIST2H3A, SPINT1, COL6A2, ZP1, CFB, ITIH4, CEMIP, CFH, C1S, C1R, CCL3L3, CFI, LYPD1, NOTCH2NL, TNFRSF11B, HPR, AOAH, LTBP3, THBS2, CFP,

Sr. No	GO terms	GO category	Gene Count	Fold Enrichment	Gene Name
					UACA, C2, SCUBE3, CCL3, CD14, APOE, CA11, GDF15, FN1, CRIM1, BMP8A, AGT, CXCL10, TG, COL5A1, ARTN, ADM2, IL17A
10	GO:0070062~extracellular exosome	CC	69	1.6	SNED1, SERPINA3, CLSTN3, ITGB3, PCDH11X, HP, OLA1, KIF12, ZDHHC1, SERPINA5, C4B, DMBT1, CDH1, C1RL, HIST1H2AG, PHGDH, PTGDS, CTSD, CAMP, SERPINB3, SLC13A3, MMP7, FBXO2, KRT34, TMC8, AZGP1, HIST2H3A, SPINT1, COL6A2, CHI3L1, SMPDL3B, UTRN, CFB, COLEC12, FBLN7, ITIH4, TNXB, CFH, C1S, C1R, CFI, SLC1A1, GLRX, LTBP2, HPR, LTBP3, KRT86, UACA, THSD4, C2, MUC1, SNX2, CD59, CD14, APOE, GDF15, CFAP70, RARRES1, FN1, CRIM1, AGT, ASS1, COL5A1, CPD, CTH, NUCB1, CPE, TPP1, FOLR1

Table 3.4: Significant KEGG pathway based on 297 differentially expressed genes.

Sr. No	Term	Gene Count	Fold Enrichment	Gene Name
1	hsa05150:Staphylococcus aureus infection	9	9.0276	C4B, CFH, C1S, C1R, CFI, CFB, HLA-DQA1, C2, HLA-DQB1
2	hsa04610:Complement and coagulation cascades	9	7.0650	C4B, CFH, C1S, C1R, CFI, CD59, CFB, SERPINA5, C2

3.2.3. Protein-Protein Interaction (PPI) network functional enrichment analysis of the 297 DEGs through Cytoscape software

PPI network was generated via STRING database using the 297 candidate DEGs. This generated PPI network was further used for identification of 208 nodes using Cytoscape software (Fig. 3a). In a PPI network, node represents the genes and the number of connections established by the nodes are denoted as 'degree' of the node. Next, STRING generated PPI network was used to identifify significant modules using the MCODE plugin of Cytoscape. The top module (module 1) consisted of 8 nodes, namely, CD59, C2, C4B, CFB, C1R, C1S, CFP and ADAMTS13 (Fig. 3b). Further, Cytoscape plugin CytoHubba was used for analysis of hub genes in order to identify genes with central functionality among the 297 DEGs. Top 8 hub genes were identified with highest maximal clique centrality (MCC) consist of the following complement genes - CFH, CFB, CFP, C4B, C1R, C1S, CD59 and C2. We also observed CFB, C2 and C4B hub genes to be a part of module 1, GO term biological processes GO:0030449 as well as these genes were common to KEGG pathways hsa04610 and hsa05150 (Fig. 3c). Therefore, for further studies we have considered CFB, C2 and C4B hub genes to be our candidate genes.

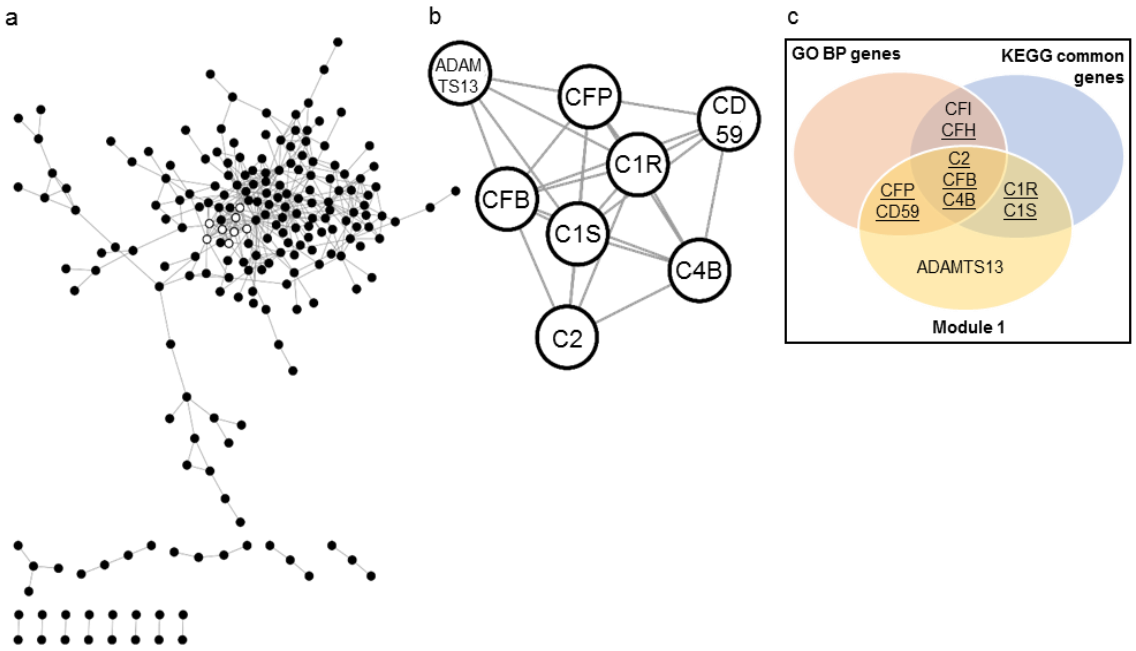

Figure 3.3. *Module analysis and identification of candidate genes. (a) PPI network of identified DEGs. (b) Module with highest score of 6.85 and consist of 8 edges. (c) Venn diagram of common genes between most enriched GO term, KEGG pathway and top module.*

3.2.4 Identification of candidate hub genes that have prognostic significance in triple negative breast cancer

In order to identify whether the candidate genes are prognostically relevant in predicting disease outcome in TNBC patients, online Kaplan–Meier plotter was used. We observed CFB (Affymetrix probe ID: 211920_at) (HR = 0.59(0.38-0.9), logrank p = 0.015) and C2 (Affymetrix probe ID: 203052_at) (HR = 0.6(0.39-0.92), logrank p = 0.019) to significantly predict relapse free survival (RFS) in TNBC patients (n=255) (Fig. 4a, 4b). Unfortunately, Kaplan–Meier plotter did not have an Affymetrix probe that was exclusive for C4B. Therefore, the C4B gene was not considered for further study. We observed that high expression of CFB and C2 complement genes increased the probability of relapse free survival in TNBC patients ($p < 0.05$).

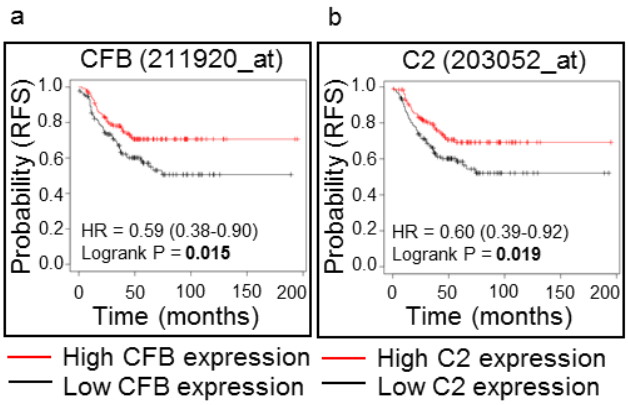

Figure 3.4. *Survival analysis of the candidate genes, (a, b) Survival analysis of CFB and C2 in TNBC patients was obtained from KM plotter (HR = 0.59(0.38-0.90), log-rank p-0.015; HR = 0.60(0.39-0.92), log-rank p- 0.019).*

3.2.5. Analysis of CFB hub genes in the context of TRAIL in triple negative breast cancer patients

To further understand the role of CFB and C2 in triple negative breast cancer in relation to TRAIL, we conducted correlation analysis through online Kaplan–Meier plotter. We observed CFB (Pearson's correlation coefficient, r = 0.13, p < 0.05) but not C2 (Pearson's correlation coefficient, r = 0.09, p > 0.05) to be positively correlated with TRAIL in these 255 TNBC patient samples (Fig. 5a). Furthermore, since CFB was observed to be correlated with TRAIL we performed the Kaplan–Meier analysis on these 255 TNBC patients after segregating them with respect to median TRAIL expression. We observed high CFB to significantly predict increased probability of RFS in high TRAIL group (HR = 0.37 (0.17-0.79), logrank p = 0.008) but not in low TRAIL group (HR = 0.8 (0.47-1.36), logrank p = 0.410) (Fig. 5b).

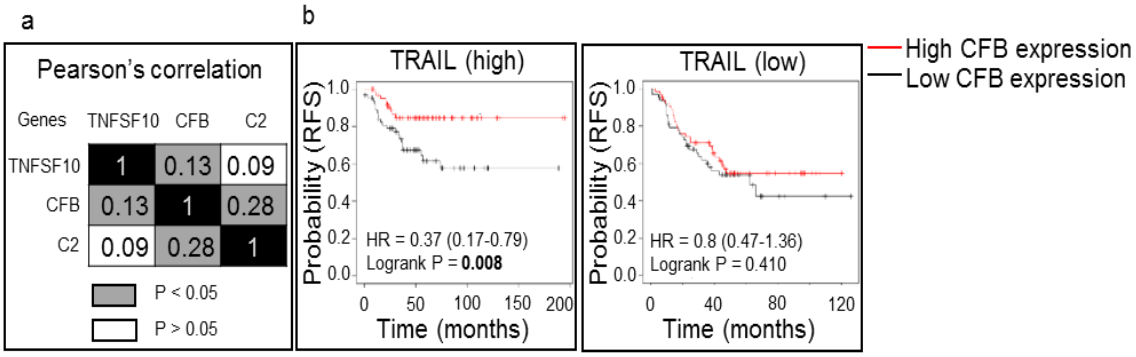

Figure 3.5. *Validation of CFB and C2 relation with TRAIL treatment. (a) Positive correlation between CFB and TRAIL expression in the TNBC (r = 0.13, p <0.05) and positive correlation between C2 and TRAIL expression in the TNBC (r =0.09, p < 0.05). (b) Survival analysis of CFB in TNBC patients divided as per high (HR = 0.37(0.17-0.79),*

log-rank p- 0.008) and low (HR = 0.8(0.47-1.36), log-rank p- 0.410) expression of TRAIL obtained from KM plotter.

3.2.6. Analysis of complement genes in the context of TRAIL treatment in MDA-MB-231 breast cancer cells

To understand the possible interactions of TRAIL with complement CFB, we used GeneMANIA software. Based on Perou et.al. 2000 and Bild et.al. 2006, GeneMANIA showed TNFSF10 to be 'co-expressed' with CFB. It also showed 'co-expression' with CD59, C1R, C2 and C1S. However, GeneMANIA did not show 'physical interaction', 'co-localization', 'gene interaction' and 'pathway' between TNFSF10 and any of the complement genes (Fig. 6a). Analysis of the microarray data that we had obtained after treating MDA-MB-231 cells with rhTRAIL showed a fold-change of the complement genes as follows: CFB (FC: 2.01), C2 (FC: 1.73), C4B (FC: 2.79), CFP (FC: 1.57), CD59 (FC: 1.86), C1R (FC: 1.76), C1S (FC: 2.12), CFI (FC: 3.32), CFH (FC: 1.80) (Figure 6b). Furthermore, we observed that the ratio of expression of genes responsible for activation (CFB, C2, C4B, CFP, C1R, C1S) the complement pathway is higher than the genes responsible for inhibiting (CD59, CFI, CFH) the complement pathway (Fig. 6b).

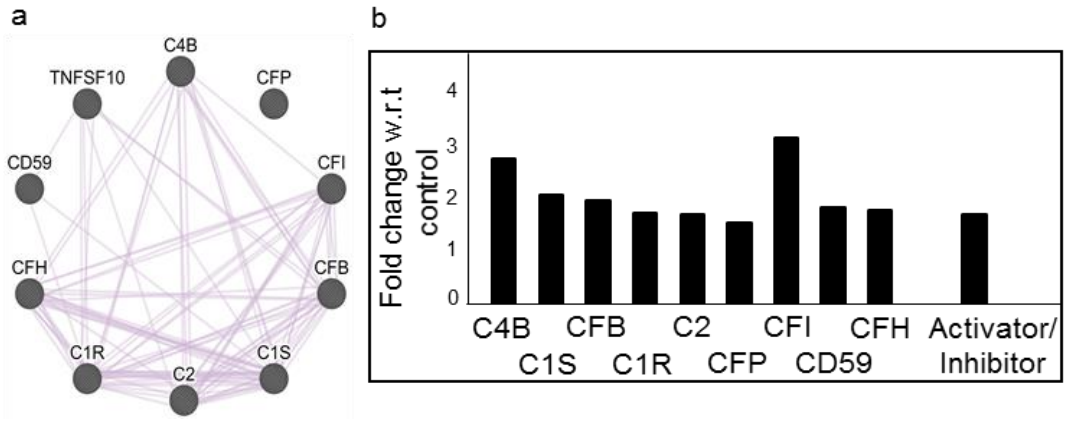

Figure 3.6*. Validation of complement activation in rhTRAIL treated MDA-MB-231 cells. (a). Gene/protein interaction networks of TRAIL and complement pathway genes was created by GeneMANIA. Violet lines represent the 'co--expressed' relation between two genes. (b) Microarray mRNA expression levels of complement pathway genes after rhTRAIL treatment and the ratio of complement activator genes vs inhibitor genes.*

3.2.7. Discussion

Breast cancer is one of the most common threats towards women's health globally. Due to poor prognostic rate, death count is increasing every year[151]. Breast have a potent heterogenetic nature, which difficult to treat even with the present advanced chemotherapy[156]. The present understanding of the biologic heterogeneity of breast cancer has help immensely in developing new drugs in recent years, but there is a search for better selective therapeutic approach. Triple negative breast cancer (TNBC) is the most aggressive form of breast cancer with very high mortality rate, as do not express estrogen receptors, progesterone receptors and Her2neu receptors[25]. The selective nature of TRAIL to induce apoptosis in the cancer cells only, made it a potential candidate for cancer treatment. Some studies have shown promising results with TRAIL in TNBC cell lines[56]. Role of TRAIL induced apoptosis in tumor microenvironment (TME) was stated in different studied. TRAIL is expressed by different immune cells in TME such as monocytes, macrophages, cytotoxic T cells and NK cells. TRAIL secreted by NK cells produces IFNγ and other cytokine to induce apoptosis in tumor cells, the amount and the half-life of TRAIL in circulation plays a crucial role in immune regulation[113,157,158]. There have been some clinical trials to inject and increase the half-life of TRAIL in circulation, one of the tested anti-tumoral agent has successfully increased TRAIL amount in MDA-MB-231 xenografts model[159,160]. As we can hypothesize that the concentration of TRAIL in TME may not be sufficient to eradicate the tumor cells, therefore in our study we have used a sub-lethal dose of TRAIL to understand regulation of different pathways under such conditions.

In our study we have treated MDA-MB-231 TRAIL sensitive cells with sublethal doses(50ng/ml) of rhTRAIL and induction of apoptosis was observed (Fig. 1a, 1b). Treated samples were used for microarray analysis and Differentially Expressed Genes (DEGs) identification. We found 297 DEGs with 200 upregulated and 97 downregulated genes (Fig. 1c, 1d, Table 3.1, 3.2). DAVID is a widely used database for annotating genes into GO Biological Processes (BP), GO Molecular Function (MF), and GO Cellular Components (CC) by accumulating data from different biological databases (NCBI, KEGG, Gene Ontology, Ensembl, Uniprot, Reactome, etc). DAVID arranges data of given genes/proteins in a specific for different functional interpretations. These TRAIL treated cells were mostly enriched in GO terms like regulation of complement activation, blood microparticle, proteinaceous extracellular matrix, extracellular matrix, extracellular space,

extracellular region, extracellular exosome, heparin binding, serine-type endopeptidase activity, and calcium ion binding. Among these significant GO terms 'regulation of complement activation' has the highest fold enrichment value of 15.1 and it comprises of 7 genes (C4B, CFH, CFI, CD59, CFP, CFB, C2). These 7 genes were found to be also present in other mentioned significant GO terms (Fig. 2a, Table. 3.3). The KEGG analysis showed pathways like Staphylococcus aureus infection and Complement and coagulation cascades with fold enrichment value of 9 and 7 respectively. The common genes between the pathways are C1R, C1S, C2, C4B, CFB, CFH, CFI. PPI network was built with these 297 DEGs and module analysis showed that the most significant module was related to Complement system and activation. The top module comprised of C59, C2, C4B, CFB, C1R, C1S, CFP and ADAMTS13. After doing the hub gene analysis, we identified CFH, CFB, CFP, C4B, C1R, C1S, CD59 and C2 with highest maximal clique centrality (MCC). We further identified the common genes i.e. CFB, C2 and C4B between the GO term 'regulation of complement activation', the common genes present between the KEGG pathways and the top module (Fig. 2b, 2c, Table. 3.4).

To further explicate the molecular mechanism underlying rhTRAIL mediated apoptosis, we constructed a PPI network for identification of connective nodes. 208 observed nodes represent respective genes in the obtained PPI network**.** This network was used to identify different protein groups or modules that participates in a particular cellular process**.** The observed primary significant module was related to complement system and activation and it contains CD59, C2, C4B, CFB, C1R, C1S, CFP and ADAMTS13 (Fig. 3a, 3b). Next, to identify the genes with highest number of interactions among each other in the given PPI network, hub gene analysis was performed. 'Degree' value of the node denotes number of interactions established by a node in the given network**.** One of the most reliable methods for assigning degrees to nodes or hub gene in a PPI network is Maximal clique centrality (MCC)**.** CFH, CFB, CFP, C4B, C1R, C1S, CD59 and C2 were observed with highest MCC degree indicating high level of interactions in the rhTRAIL activated signalling pathway. Interestingly we found that all the three previously identified common genes (CFB, C2 and C4B) are also among the top hub genes and they are all part of the complement cascade. After doing survival analysis for prognostic value of these genes, we have identified two genes (CFB and C2) with positive prognostic values.

Complement system is a key factor in our bodies first line of defence against the foreign invasion, it comprises of a cascaded of serine proteases. Components of complement

system provides a platform for cross-talk between complement system and systemic and regulatory functions. Complement system as a definitive role in innate immune system, adaptive immune system, organ development, homeostasis, regulation of the coagulation system, synaptic maturation, angiogenesis, mobilization of hematopoietic stem-progenitor cells, tissue regeneration, lipid metabolism and even in cancer regulation. Complement system can be activated by either of these 3 well known pathways: Classical pathway or the antigen-antibody complex formation system; Lectin pathway and Alternative pathway. It plays an important role in tumor immune surveillance but sufficient studies are not present to give a clear picture of how complement system regulate tumor survival, growth and progression. complement factor b (CFB), an integral part of the alternate pathway of complement system is known biomarker for pancreatic ductal carcinoma. Suman et al. have shown that CFB was modulated in all the subtypes of breast cancer in both serum and plasma level. They suggested that CFB could be considered as a prognostic marker for luminal-A subtype of breast cancer. CFB was also identified as a prognostic marker for pancreatic cancer and it is a favourable prognostic marker for breast cancer is mentioned in Human protein atlas. On the other hand C2(Complement Component 2) which is part of the classical and lectin pathway of complement cascade, have been identified as a favourable prognostic marker for HCC. Even Single Nucleotide Polymorphism (SNP) were also identified in relation with risk of developing HCC.

Kaplan-Meiers survival analysis identified increased expression of CFB and C2 genes to show high relapse-free survival in the 255 TNBC patients (Fig. 4a, 4b). Thus, CFB and C2 genes that are upregulated on exogenous rhTRAIL treatment in MDA-MB-231 cells, showed better outcome in survival analysis for TNBC patients. This indicated increased CFB and C2 expression to be responsible for better prognosis in TNBC patients. Additionally, positive correlation was observed between TRAIL and CFB transcript expression in the analysed TNBC patients (Fig. 5a). On stratification of TNBC patients as per the endogenous TRAIL expression we observed that in high TRAIL expressing group, high CFB cohort showed better relapse free survival compared to low CFB cohort (Fig. 5b). Similar to TRAIL, CFB is also known to induce caspase-3 mediated apoptosis in cancer cells. This indicates that the patients with simultaneous upregulation of proapoptotic proteins, TRAIL and CFB, have a better outcome in the context of developing secondary tumor over a period of 200 months. Taken together, our results suggest that treating patients with TRAIL might help in increasing the CFB-mediated complement

response against tumor. Interestingly, we observed rhTRAIL resistant MDA-MB-231 cells to show decreased expression of CFB (unpublished data) against its increased expression in rhTRAIL sensitive MDA-MB-231 cells. Taken together, we can conclude that alternative pathway of the complement system especially CFB plays an important role in enhancing the sensitivity of triple negative breast cancer cells towards TRAIL treatment.

Corelated gene/protein analysis through GeneMANIA of TRAIL and complement related genes from our overall study (CFB, C2, C1S, C1R, C4B, CFP, CFH, CFI and CD59) showed that TNFSF10(TRAIL) 'co-expression' with CFB, C2, CD59, C1S and C1R. From the microarray data of rhTRAIL treated MDA-MB-231 cells, we found that all the 9 complement related genes are upregulated and the observed ratio value of genes responsible for activating the pathway are greater than those genes which are responsible for inhibition of the complement pathway. This result indicates that there is a high chance of activation of complement pathway via TRAIL treatment (Fig. 6a, 6b). In the tumor microenvironment complement pathway could be activated via both classical and alternative pathways. rhTRAIL treatment upregulates activators of both the pathways like C1R,C1S, C2, C4b of classical pathway and CFB, CFP of alternative pathway, but in contrast it also upregulates the inhibitors like CFH,CFI and CD59(Fig.7). As stated earlier the ratio of activator genes is higher than the inhibitor genes, we can hypothesize that TRAIL is responsible for activation of the complement pathway in TME of triple negative tumors and enhanced expression of CFB and C2 will result in better RSF of TNBC patients.

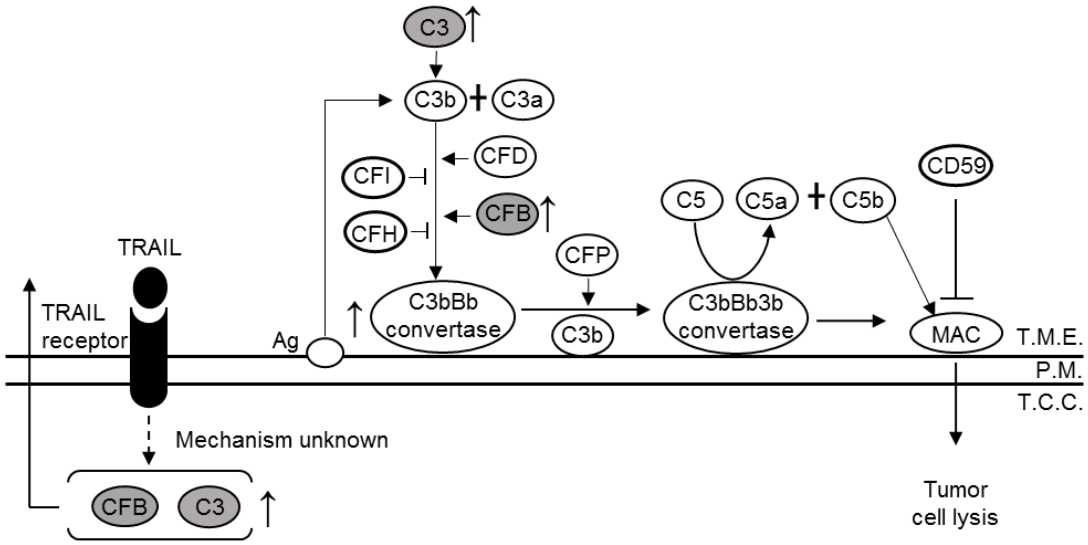

Figure 3.7. *Schematic representation of the complement pathway genes influenced by TRAIL. TRAIL after binding with the death receptors by an unknown mechanism*

upregulates few of the complement pathway genes. In classical pathway TRAIL upregulates C1r and C1s, where C1r present on C1q cleaves and activates C1s which in turns cleaves upregulated C2 and C4 into C2a and C4b, and form the C3 convertase. In case of alternative pathway TRAIL upregulates CFB and CFP, which generates the C3bBb convertase that leads to the formation of C5 convertase and then MAC formation. MAC facilitates cell lysis and apoptosis. TRAIL also upregulates CFH, which gives hindrance to the C3 convertases of classical pathway and the process of formation of C3bBb3b convertases in alternative pathway. TRAIL increases CFI inhibitor of C3 convertases of alterative pathway and CD59, which inhibits MAC formation. The calculated ratio of activator genes vs inhibitor genes is found to be high which indicates increase of cell lysis and apoptosis. Deep gray represents activator genes and light gray represents inhibitor genes in both the pathways.

TRAIL is a protein that has been known to initiate apoptosis in cancer cells without targeting normal cells. Our study for the first time shows TRAIL to be capable of activating the complement pathway of innate immunity in TNBC cells. We further identified complement factor B (CFB) to have a favorable prognostic value in triple negative breast cancer patient expressing high levels of endogenous TRAIL. Thus, our study indicates the complement system to be involved in promoting TRAIL mediated cell death which needs to be explored further in the context of triple negative breast cancer.