ISSN: 2755-1067

Journal of Diseases Disorders & Treatments

SCIENTIFIC Research and Community

Research Article

Open Access

Impact of Tumour Genomics on Response to Standard Neo-adjuvant Chemotherapy in Triple-Negative Breast Cancer

Divya Dhananjay Singh[1]* and Aditya Bardia[2]

[1]Masters Student Year II at MMSCI Program, Harvard Medical School, USA

[2]MGH Cancer Centre, Massachusetts General Hospital, USA

SUMMARY

Background: The genomic delineation of breast cancer is a novel approach to evaluating the disease. The opportunities arising from this genomic data analysis help in identifying specific cell line models and subtypes that are key for preclinical studies, developing predictive biomarkers, new target agents, and informing therapy selection. There is a wealth of information in previously collected datasets from breast cancer patients, but this has been under-studied and information on clinical association from these datasets has been limited. Triple-Negative Breast Cancer (TNBC) has been especially notorious since a large number of young premenopausal women diagnosed with the disease have suffered a worse outcome than the other known breast cancer types and there is no targeted and effective therapy for the treatment of the disease. This approach of taking a closer look at the genomics of TNBC will help in moving toward the often-cited dream of personalized medicine in difficult to treat areas of breast cancer, more specifically in our case, Triple Negative Breast Disease. However, the adaptation of genomic data into treatment regimens customized or tailor-made for each individual patient has been restricted in part, due to uncertainties in diverging from guideline-based clinical protocols.

We report the Genomic Approach to TNBC in the context of predicting favorable outcomes to treatment responses and briefly describe recurrence-related data.

Patients and Methods: We studied data derived from three different gene expression profiling platforms. We identified 486 samples associated with disease response to therapy and selected them for in-depth analyses. Hence 3 independent sets of analyses were carried out for each different platform due to the fact that each of these platforms had gene expression profiles that were populated with the TNBC cases with specific neo-adjuvant therapies namely, Platinum-based cohort [referring to HGU133plus2 platform related samples], taxane cohort [referring HGU133A platform samples] and anthracyclines with or without taxane cohort [HTA2.0 array] respectively.
We aimed to analyze recurrence data, however, it was sparse amounting to 50 samples in total across the various publicly available datasets and we could analyze the same only but could not make definitive conclusions owing to the small sample size.

Results: In the platinum-based cohort, pathway enrichment analysis using Cluster Profiler revealed the enrichment of key terms like Platinum drug resistance, chemical carcinogenesis, glutathione metabolism-related terms in no response category. In the Taxane-based cohort, pathway enrichment analysis revealed a defined enrichment of oncogenic signaling pathways like Pi3K-AKT, ECM, Focal adhesion in the no response category and the Anthracycline with or without taxane cohort showed that while enrichment of Wnt and TGF beta signaling were observed in no response category the response group was seen populated with immune response-related pathways.

Conclusions: Chemotherapy is an important treatment option and is currently the mainstay in cases of Triple-Negative Breast Cancer. It was shown to reduce one-third of the annual death rate regardless of tumour characteristics. Chemotherapy is used as the mainstay in cases of triple-negative disease, HER2-positive tumors and high-risk oestrogen positive tumors. Significant toxicity associated with the therapy aside, not every patient benefits compulsorily from it and there is a sheer dearth of both practically validated and theoretically proven predictive markers to allow the tailoring of chemotherapy regimens to each individual patients needs or tumour type. Achieving pathological complete response (pCR) is associated with favorable outcomes, however, its predictive power in being used as a surrogate endpoint for improved overall survival (OS) has been debated. TNBC and breast cancer in general, need to be viewed in the context of underlying genomic phenotypes. Based on this notion, treatment decisions should be less influenced by guidelines based on set clinical protocols and rather be more driven by tumour biology if we are to improve the response to chemotherapy.

***Corresponding author**
Divya Dhananjay Singh, Masters Student Year II at MMSCI Program, Harvard Medical School, USA. E-mail: singhdivya66@gmail.com

Received: June 22, 2022; **Accepted:** June 24, 2022; **Published:** July 04, 2022

Introduction

Breast cancer is the top-ranking malignancy in women and has officially surpassed lung cancer as the most commonly occurring carcinoma worldwide in 2020 [1]. Except for triple negative breast cancer (TNBC) the 5-year survival is typically >90%. TNBC which is defined by Estrogen Receptor, Progesterone Receptor, and HER2/ERBB2-negativity and is a clinically defined subgroup of breast cancer, constituting approximately 10-15% of cancers in western nations. TNBC is over-represented among younger women, premenopausal women, African American women, and women with inherited mutations in high-penetrance breast cancer susceptibility genes [2,3]. TNBC tumors are highly heterogeneous on the molecular level, involving differences in genetic features, germline alterations (e.g., BRCA1 mutations), DNA repair

deficiency, epigenetic alterations, gene expression patterns, and morphological features [3-9]. Several of these features have been associated with prognosis and/or therapy response in TNBC patients. Chemotherapy and surgery were the two most important treatment options available to both patients and oncologists and both have proved to be instrumental in decreasing the mortality rates of breast cancers significantly [10]. However, adjuvant and Neo-adjuvant chemotherapy has shown to have poor response rates in Triple negative breast cancer. It has been proven to be less efficacious. The need to identify which patients will actually derive benefit from chemotherapy is a crucial task and recent studies infer, in various types of cancers, genetic screening can be a encouraging tool to facilitate personalized chemotherapy regimens of individual cancer patients [11-13].

Despite the significant work and scientific advances made in the field, there have been limitations when it comes to application of these principles to real world patients. Firstly, the genetic testing systems have not demonstrated the capability to predict treatment responses to the standard Neo-adjuvant Chemotherapy and therefore do not assist in the selection of specific treatment regimens for each individual patient [14]. Secondly, most testing systems apprise us on cell proliferation events, while other mechanism-related information and therapeutically relevant findings are not fully considered, especially information on the modulation of drug targets for breast cancer therapies [15]. Conventionally, the assessment of the effectiveness of a given chemotherapy has been based on the the pathologic complete response (pCR). This term is defined by "the complete lack of signs for cancer relapse after radiation and/or chemotherapy." Prediction of pCR based on genetic screening of individual patients has not been initiated anywhere nor as mentioned previously can pCR serves as a surrogate endpoint for improved overall survival [16]. Our study aims to develop a gene expression signatures that can be used to identify TNBC patients with improved responses to chemotherapy. These would be a combined group of genes within a cell with a uniquely characteristic pattern of expression that occurs as a result of an altered biological processes. This involves analyzing through thousands of genes or genetic data derived from high throughput genetic technologies or molecular libraries and selecting a few that will predict the outcome. However, the accuracy of these predictions for each signature needs validation otherwise the results will be biased.

As a proof-of-concept for genome-guided chemotherapy, predicting response versus no response for chemotherapy, we performed analysis on around 500 breast cancers gene expression profiles from three different independent publicly available databases and identified genes which were associated with chemotherapy response in breast cancers. We could not quantify the risk of recurrence or draw inferences from the data as the sample size was approximately 50 (extremely small).

Materials and Methods

Datasets

TNBC data collection and processing: Gene expression profiles of TNBC patients with information about therapy response and specifically treated with neoadjuvant therapies were collected from GEO. Data was derived from the NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). Totally 500 samples from 3 different gene expression profiling platforms like Affymetrix, HGU133plus 2 array [Total number of samples: 45], HGU133A array [Total number of samples: 322], and HTA2 array [Total number of samples: 119], were collected and used for the study. After the removal of duplicates 486 samples were obtained.

We collected 500 samples, in total, from 3 different gene expression profiling platforms like Affymetrix, HGU133plus 2 array [Total number of samples: 45], HGU133A array [Total number of samples: 322], and HTA2 array [Total number of samples: 119], were collected and used for the study.

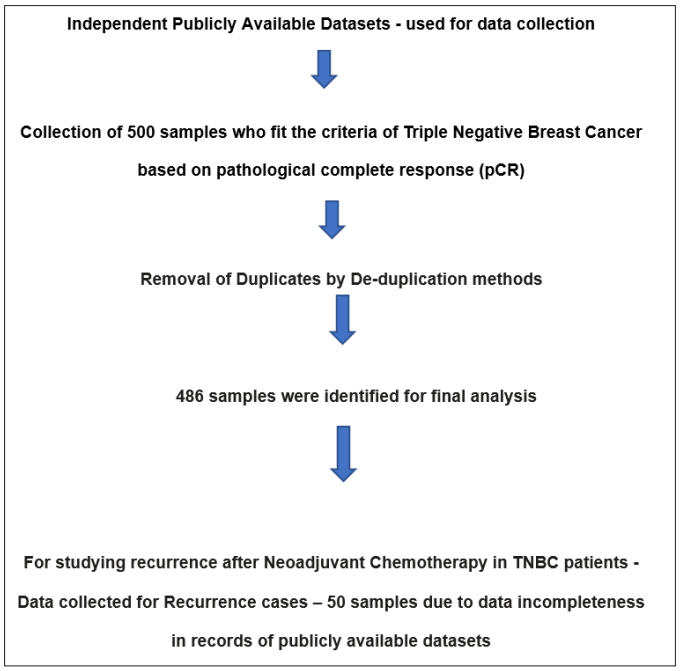
After the removal of duplicates 486 samples were obtained. "Duplication" means that the publicly available datasets used for data collection may have repeated data. This could be due to clerical errors such as data entry errors or data collection methods. For example, while using a web scraper it may happen that one scrapes the same webpage more than once, or the same information from two different pages. Duplication can lead us to make incorrect conclusions by leading us to believe that some observations are more common than they really are.

Removing duplicates is called "deduplication" These are done using the following techniques in data analysis software and form a part of data cleaning. It is a critical step before using the collected data for final analysis.

- Visualizing duplication
- Finding & removing exact duplicates
- Finding & removing partial duplicates

The 50 samples, on the other hand, were the recurrence-related data which were available for only 50 cases. Cancer recurrence is a critically important outcome for patients and clinicians. However, publicly available cancer registry data rarely contain recurrence information. The National Cancer Data Base (NCDB) collects recurrence data; however, this information is not provided to researchers because of completeness and accuracy concerns.

Also, the majority of hospitals report incomplete recurrence information for more than half of their patients. The presence of incomplete recurrence information is largely dependent on undefined hospital factors, rather than patient or tumour characteristics. Attempts to improve cancer recurrence information should focus on hospital operational and process factors surrounding how the hospital tumour registries collect recurrence data and make it available to the research community for analysis.



CONSORT Diagram Showing Flow of Data Collection Methods

It is a database of gene expression curated profiles maintained by the National Centre for Biotechnology Information (NCBI) and is called the Gene Expression Omnibus (GEO). It provides access to all its data sets and outputs a data set selected for further processing. It is a curated data set derived from investigator-submitted data.

GEO (Gene Expression Omnibus) is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community. GEO data can be searched through the GEO Datasets and the GEO Profiles databases.

The three main goals of GEO are to:

1. Provide a database of high-throughput functional genomic data
2. Support complete and well-annotated data deposits from the research community
3. Allow users to query, locate, review and download studies and gene expression profiles of interest

The Gene Expression Omnibus (GEO) at the National Centre for Biotechnology Information (NCBI) is also the largest fully public repository for high-throughput molecular abundance data, primarily gene expression data. The database has a flexible and open design that allows the submission, storage and retrieval of many data types. These data include microarray-based experiments measuring the abundance of mRNA, genomic DNA, and protein molecules, as well as non-array based technologies such as serial analysis of gene expression (SAGE) and mass spectrometry proteomics technology. GEO currently contains over 30000 submissions representing approximately half a billion individual molecular abundance measurements, for over 100 organisms. The GEO database is publicly accessible through the World Wide Web at http://www.ncbi. nlm.nih.gov/geo

Hence 3 independent sets of analysis, each for the different platforms were performed. As these platforms had gene expression profiles that were populated with the TNBC cases with specific neo-adjuvant therapies to Platinum-based [referring to HGU133plus2 platform related samples], taxane cohort [referring HGU133A platform samples] and anthracycline with or without taxane cohort [HTA2.0 array] respectively.

The raw (.cel files) used in the respective cohort were downloaded from GEO. The raw .cel files of the independent cohorts were Robust MultiArray Average (RMA) normalized using Genespring GX and Pathway Finder software (Agilent Technologies, USA). As the anthracyclines with or without taxanes cohort were from a single GSE profile (GSE106977) we used the RMA normalized values from GEO as obtained.

Statistical Analysis

The normalized gene expression measure of 1) platinum-based, 2) taxane-based and (3) anthracycline with or without taxane based cohort were used for differential gene expression analysis. Genes with 1.5 fold differential expression among response Vs no response category with p-value<0.05 were considered significant. The enrichment analysis of the genes was performed using Cluster Profiler.

When we test for Differential Expression of Genes (DE) we start by assuming that the tested treatment has no effect (null hypothesis). The p-value is the probability of obtaining the observed or more extreme result if the null hypothesis were true. A small p-value means the null hypothesis is probably not true and the treatment does have an effect.

A large p-value indicates that the null hypothesis is plausible, still, the treatment could have an effect. The null hypothesis is never true. If we can increase the sample size we can get p-values as small as we want. So we can identify which genes are affected but can't say which genes are not affected. Small p-values are usually correlated to logFC. However, genes with many read counts have higher statistical power to detect DE, in these cases we can see small p-values for genes with logFC close to zero.

In practice, to get DE genes we filter for False Discovery Rate (an adjusted p-value) less than, say, 0.05. Optionally, we filter also for absolute logFC above a certain threshold (depending on the biological question). It usually doesn't make sense to invest time in genes with logFC close to 0 even if the p-value is very small. To get genes non-DE we could filter for logFC close to zero and p-value above, say 0.2, but the interpretation of p-value from above explanation needs to be understood. (Bayesian statistics suffer less of this "issue").

To select covariates for each case study, we visually evaluated whether each covariate was informative by examining a scatter plot of the independent covariate percentile and the p-value. If this contained any noticeable trend such that certain values of the informative covariate were enriched for smaller p values, we considered the covariate to be informative.

Most classically, methods that control the family-wise error rate (FWER), or probability of at least one false discovery, have been used to correct for multiple testing. Despite their popularity, FWER-controlling methods are often highly conservative, controlling the probability of any false positives (type I errors) at the cost of greatly reduced power to detect true positives. The trade-off of type I errors and power becomes exacerbated in the analysis of data from high-throughput experiments, where the number of tests being considered can range from several thousand to several million.

The Benjamini and Hochberg step-up procedure (BH) is the method we used to control the FDR.

Note on Missing Data - There was no missing data as we used the raw files (.cel files) and these were Robust Multiarray Average (RMA) Normalized before final analysis.
https://felixfan.github.io/RMA-Normalization-Microarray/

The detailed statistical analysis is as follows:

1. Differential gene expression analysis in Genespring GX (Version 14.9) was done using the unpaired t-test. GeneSpring GX https://www.agilent.com/en/product/software-informatics/genomics-software-informatics/gene-expression/genespring-gx
2. For p-value calculation, fold change 1.5 above or -1.5 below was used.
3. Benjamin Hochberg correction-based False Discovery Rate was performed to calculate the adjusted p-value.
4. Clusterprofiler (Version 3.18.1) calculates the hypergeometric distribution based p-value in its background which is used for the plots. Clusterprofiler https://bioconductor.org/packages/

release/bioc/html/clusterProfiler.html

5. Hierarchical clustering based on Pearson correlation measures of gene expression pattern was used in heatmap visualization.
6. Volcano plot and heatmaps were performed using Complex Heatmap (Version 2.6.2) and ggplot2 packages of R (version 4.0.5). ComplexHeatMap - https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html Ggplot2 package - https://r-graph-gallery.com/ggplot2-package.html
7. Scripts used for the heatmaps, volcano plots provided in Additional Files.

Results

Heterogeneous treatment response to therapies in TNBC:

In this study we focussed on genomic impact of neo-adjuvant therapy response in TNBC patients. Out of the 483 samples analysed from 3 different cohort, the detailed report on the results are provided below.

1) Platinum-based cohort: Transcriptome profile of neoadjuvant trial from cisplatin monotherapy or platinum & bevacizumab treated TNBC patients dominate this cohort. Differential gene expression analysis of treatment response to no response category in this cohort resulted in upregulated of 91 genes and downregulation of 227 genes with significant p-value <0.05. Volcano plot & Heat map depicting the expression pattern of genes differentially regulated across noresponse and response category are shown in Figure 1. Further pathway enrichment analysis using cluster profiler revealed the enrichment of key terms like Platinum drug resistance, chemical carcinogenesis, gluatathione metabolism related terms in no response category (Figure 1).

2) Taxane-based cohort: Neoadjuvant therapy of taxane-antracycline chemotherapy or Taxol based drugs along with FEC dominate (total number of samples: 322) these cohort. Differential gene expression analysis of treatment response to no response category in these taxane-based cohort resulted in the upregulated of 12 genes and downregulation of 30 genes in the no response category. Volcanoplot and Heatmap showing the expression pattern of differential expressed genes was shown in Figure 2. Pathway enrichment analysis revealed a defined enrichment of oncogenic signalling pathways like Pi3K-AKT, ECM, Focal adhesion etc., in the noresponse category.

3) Anthracycline±taxane cohort: Totally 119 samples, with 88 patients treated with neoadjuvant anthracyclines and/or taxanes and 31 with antracyclines and/or taxane plus carboplatin constituted this cohort. Analysis of this cohort resulted in 10 upregulated and 47 downregulated genes in no-response category. The corresponding volcanoplot and heatmap showing their expression pattern are provided in the Figure 3. While enrichment of Wnt and TGF beta signalling were observed in noresponse category the response group was seen populated with immune response related pathways. Though the enrichment results were not highly significant as the total number of differentially regulated genes were very less, it provide us an idea on the action of immune response pathways in conferring better immune response to TNBC patients Figure 4-13.

Baseline Characteristics Table: Table showing baseline characteristics (sex,race,age,region) from datasets

	Total cancer number	Nontriple-negative number (%)			Triple-negative number (%)
		HR+ Her2−	HR+ Her2+	HR− Her2+	
Sex					
Female	295,801	214,052 (72.4%)	29,794 (10.1%)	13,327 (4.5%)	38,628 (13.1%)
Male	3,136	2587 (82.5%)	315 (10.0%)	49 (1.6%)	185 (5.9%)
Race/ethnicity					
Non-Hispanic white	235,082	175,760 (74.8%)	22,870 (9.7%)	9669 (4.1%).1%)	26,783 (11.4%)
Non-Hispanic black	33,970	20,255 (59.6%)	3744 (11.0%)	1904 (5.6%)	8,067 (23.7%)
Non-Hisp Asian/P.I.	9,294	6519 (70.1%)	1091 (11.7%)	639 (6.9%)	1,045 (11.2%)
Hispanic	15,536	10,476 (67.4%)	1847 (11.9%)	907 (5.8%)	2,306 (14.8%)
Other/unknown	5,055	3629 (71.8%)	557 (11.0%)	257 (5.1%)	612 (12.1%)
Age					
≤30	2,059	1014 (49.2%)	411 (20.0%)	155 (7.5%)	479 (23.3%)
31–40	15,094	8439 (55.9%)	2494 (16.5%)	978 (6.5%)	3,183 (21.1%)
41–50	51,793	34,894 (67.4%)	6422 (12.4%)	2627 (5.1%)	7,850 (15.2%)
51–60	72,543	49,920 (68.8%)	8051 (11.1%)	4149 (5.7%)	10,423 (14.4%)
61–70	77,870	59,173 (76.0%)	6724 (8.6%)	3061 (3.9%)	8,912 (11.4%)
>70	79,578	63,199 (79.4%)	6007 (7.5%)	2406 (3.0%)	7,966 (10%)
Geographic region					
East South Central	17,319	11,844 (68.4%)	1850 (10.7%)	888 (5.1%)	2,737 (15.8%)
West South Central	23,951	16,634 (69.5%)	2693 (11.2%)	1150 (4.8%)	3,474 (14.5%)
South Atlantic	65,180	46,520 (71.4%)	6502 (10.0%)	3035 (4.7%)	9,123 (14.0%)
East North Central	52,707	38,085 (72.3%)	5162 (9.8%)	2301 (4.4%)	7,159 (13.6%)
Middle Atlantic	46,259	33,908 (73.3%)	4699 (10.2%)	1982 (4.3%)	5,670 (12.3%)
West North Central	21,923	16,223 (74.0%)	2121 (9.7%)	949 (4.3%)	2,630 (12.0%)
Mountain	14,382	10,682 (74.3%)	1377 (9.6%)	619 (4.3%)	1,704 (11.8%)
Pacific	38,647	28,794 (74.5%)	3842 (9.9%)	1699 (4.4%)	4,312 (11.2%)
New England	18,569	13,949 (75.1%)	1863 (10.0%)	753 (4.1%)	2,004 (10.8%)

Distribution of Race/Ethnicities across the cases in datasets

	Total cancer number	Nontriple-negative number (%)			Triple-negative number (%)
		HR+ Her2−	HR+ Her2+	HR− Her2+	
White	249,185	185,265 (74.3%)	24,556 (9.9%)	10,497 (4.2%)	28,867 (11.6%)
Black	34,380	20,529 (59.7%)	3790 (11.0%)	1926 (5.6%)	8,135 (23.7%)
Asian Indian	576	383 (66.5%)	61 (10.6%)	42 (7.3%)	90 (15.6%)
American Indian/Eskimo	808	548 (67.8%)	95 (11.8%)	47 (5.8%)	118 (14.6%)
Korean	679	435 (64.1%)	90 (13.3)	62 (9.1%)	92 (13.5%)
Asian Indian or Pakistani	849	592 (69.7%)	100 (11.8%)	55 (6.5%)	102 (12.0%)
Chinese	1,448	1003 (69.3%)	167 (11.5)	105 (7.3%)	173 (11.9%)
Vietnamese	497	325 (65.4%)	63 (12.7%)	52 (10.5%)	57 (11.5%)
Japanese	903	696 (77.1%)	74 (8.2%)	31 (3.4%)	102 (11.3%)
Pacific Islander, NOS	576	419 (72.7%)	66 (11.5%)	29 (5.0%)	62 (10.8%)
Other Asian, NOS	2,266	1621 (71.5%)	253 (11.2%)	157 (6.9%)	235 (10.4%)
Filipino	1,579	1102 (69.8%)	226 (14.3%)	111 (7.0%)	140 (8.9%)
Other	2,030	1396 (68.8%)	248 (12.2%)	112 (5.5%)	274 (13.5%)
Unknown	3,161	2325 (73.6%)	320 (10.1%)	150 (4.7%)	366 (11.6%)

Table 1: shows the source of samples from different independent datasets

Platform used	GEO number	Country	Sample size (TNBC)	Subject
GPL570	GSE18864	Denmark	24	Neoadjuvant trial of cisplatin monotherapy
	GSE103668	Denmark	21	Neoadjuvant platinum & bevacizumab treated triple negative breast cancer patients
	GSE25055	USA	114	Neoadjuvant taxane-anthracycline chemotherapy in breast cancer
	GSE25065	USA	64	Neoadjuvant taxane-anthracycline chemotherapy in breast cancer
	GSE20271	USA	58	Response to different chemotherapy treatments (FEC, FAC and/or Taxol
	GSE20194	China	62	TFAC /FEC or FAC response
GPL96	GSE42822	USA	24	Neoadjuvant chemotherapy with four cycles of 5-fluorouracil/epirubicin/cyclophosphamide (FEC) followed by four cycles of docetaxel/capecitabine (TX) on US Oncology clinical trial 02-103.
HTA2	GSE106977	Spain	119	88 patients treated with neoadjuvant anthracyclines and/or taxanes and 31 with anthracyclines and/or taxanes plus carboplatin

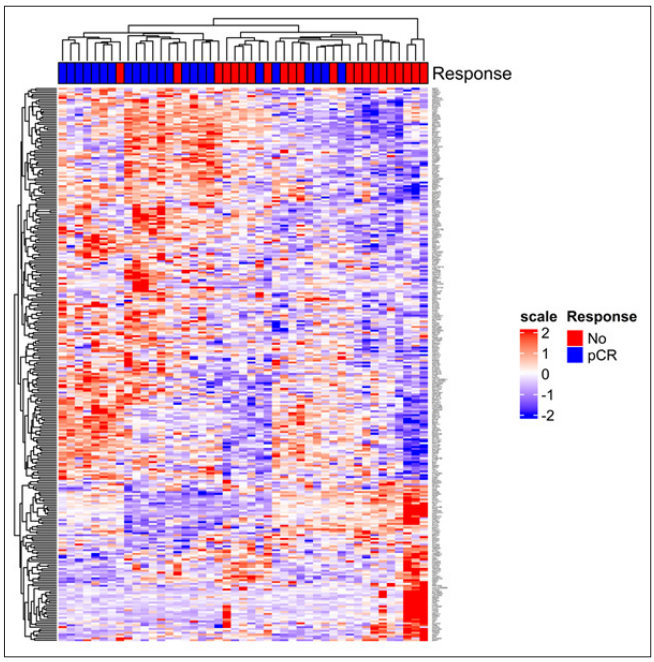


Figure 1: The Heatmap below shows the hierarchical Clustering pattern and expression of differentially expressed genes across response and no response categories of patients in GPL570 profile samples. Each column in the heatmap denotes the samples and each row denotes the expression of genes. Red colour denotes higher expression and blue colour denotes the lesser expression of genes. The red and blue lines above the heatmap gives sample information: Red -No Response Category and Blue - Pathological Complete Response Category of Patients

Table 2: GPL570 Samples - This is the list of 91 upregulated genes and 227 downregulated genes in No response category

ID	log2FC	pvalue	ID	log2FC	pvalue	ID	log2FC	pvalue
CT83	-1.83	5.03E-03	E2F7	-0.83	1.86E-02	ECT2	-0.68	4.98E-02
PSAT1	-1.60	7.29E-04	NCAM1	-0.83	1.91E-02	LINC00622	-0.68	2.89E-02
ZIC1	-1.57	5.71E-03	CHML	-0.83	1.11E-02	RNF144B	-0.68	6.87E-03
SLC6A14	-1.55	1.29E-02	BCL11A	-0.82	3.17E-03	QSOX1	-0.67	2.96E-02
MMP1	-1.42	3.22E-02	NRG1	-0.82	1.26E-02	UHRF1	-0.67	6.45E-03
MFI2	-1.34	5.72E-03	NOL4L	-0.82	1.70E-03	LINC00662	-0.67	2.62E-03
MFSD4	-1.31	2.00E-03	TMEM108	-0.82	2.02E-02	ABCD3	-0.67	2.87E-02
LOC728715	-1.28	1.02E-02	CTSV	-0.82	3.90E-03	SLC6A8	-0.67	4.65E-02
SCUBE3	-1.24	3.34E-02	KRT4	-0.82	4.25E-02	UBXN7	-0.67	3.44E-02
S100P	-1.24	3.12E-02	RNF183	-0.82	2.83E-02	CNTNAP3	-0.67	3.13E-02
LINC01133	-1.23	4.33E-03	BLM	-0.81	1.48E-03	STOX2	-0.67	3.78E-02
LINC00511	-1.23	8.67E-04	MAP3K13	-0.81	1.16E-02	ZAK	-0.67	5.18E-03
FREM2	-1.23	3.16E-03	GPR143	-0.80	3.87E-02	ZNF850	-0.66	1.58E-02
CLDN10	-1.18	3.56E-02	LAMC2	-0.80	2.78E-02	GDPD1	-0.66	3.62E-02
UGT8	-1.18	1.47E-02	PMAIP1	-0.80	1.64E-02	MALSU1	-0.66	2.04E-04
RNF182	-1.13	1.90E-02	LYPD6B	-0.79	2.76E-02	CEP55	-0.66	4.60E-02
SOAT1	-1.12	4.50E-03	GPC4	-0.79	2.01E-02	MED21	-0.66	8.47E-03
KRT16	-1.10	4.82E-02	BRIP1	-0.79	2.42E-02	CADM1	-0.66	2.72E-02
ELAVL2	-1.09	4.79E-02	STK26	-0.79	1.86E-02	CDC42EP3	-0.66	1.09E-02
LGALS7	-1.09	4.64E-03	BUB1	-0.79	4.38E-02	RNF138	-0.65	5.80E-03
ABCA13	-1.08	7.47E-03	RSAD2	-0.79	3.44E-02	KIAA1804	-0.65	2.98E-02
LEMD1	-1.07	3.56E-03	KCNQ5	-0.78	4.87E-02	RAB39B	-0.65	1.92E-02
CDCA7	-1.06	9.95E-03	ZFP42	-0.77	3.17E-02	FAM20B	-0.65	4.05E-02
MIA	-1.06	3.90E-02	MAP7D2	-0.77	3.95E-02	RQCD1	-0.65	3.22E-02
FUT9	-1.05	2.22E-02	PSIP1	-0.77	9.60E-03	ADGRL1	-0.65	3.55E-02
ELOVL6	-1.05	4.33E-03	TMEM170B	-0.77	9.09E-04	B3GALNT1	-0.65	2.24E-02
IGFBPL1	-1.04	5.51E-03	CEMIP	-0.77	2.36E-02	UNK	-0.65	3.10E-02
CXCL1	-1.04	1.98E-02	ARNTL2	-0.77	6.09E-03	INA	-0.65	4.03E-02
TMEM139	-1.02	1.19E-02	FRAS1	-0.76	4.25E-02	IL17RD	-0.65	4.51E-02
IDO1	-1.01	1.15E-02	PREP	-0.76	3.55E-03	FOLH1B	-0.65	2.24E-02
SHROOM2	-1.00	4.56E-03	CDC6	-0.76	8.50E-04	RAB38	-0.64	4.08E-02
IGF2BP2	-0.99	2.64E-02	RIF1	-0.76	4.13E-02	SAP30BP	-0.64	2.14E-04
DYNC1I1	-0.99	1.28E-02	SBK1	-0.75	9.03E-03	MAGOHB	-0.64	2.47E-02
ARNT2	-0.98	1.35E-03	CYP27C1	-0.75	7.24E-03	ABI2	-0.64	3.60E-02
SUSD4	-0.98	5.35E-03	TUBGCP3	-0.75	5.73E-03	SPTSSB	-0.64	1.71E-02
ZNF711	-0.97	9.34E-03	SHC3	-0.75	1.62E-02	MTAP	-0.64	3.66E-02
FAM84A	-0.97	9.52E-03	SSFA2	-0.74	2.41E-03	FXR1	-0.64	1.62E-02
B3GNT5	-0.97	1.95E-03	RAB36	-0.74	9.28E-04	SYN2	-0.63	5.64E-03
BCL2L14	-0.97	1.18E-02	XK	-0.74	4.99E-02	NGB	-0.63	2.62E-02
SLC35F3	-0.97	1.52E-03	NAP1L5	-0.73	2.46E-02	SMC4	-0.63	1.57E-02
PCDH8	-0.96	2.33E-02	SPIN4	-0.73	4.14E-02	BRWD1	-0.63	2.22E-02
HS6ST2	-0.96	4.36E-02	ENPP4	-0.73	4.89E-03	BACE2	-0.63	4.22E-02
EPHA7	-0.95	2.93E-02	CXCL8	-0.72	4.80E-02	MYO10	-0.63	3.60E-02
AIM2	-0.94	1.62E-02	WDHD1	-0.72	2.92E-02	TSPAN12	-0.63	2.43E-02
C6orf141	-0.94	2.35E-02	PKIA	-0.72	7.61E-03	ABCA17P	-0.63	3.78E-02
SIX3	-0.94	3.33E-02	KIF18A	-0.72	1.19E-02	SSR3	-0.63	2.36E-02
CXCL6	-0.93	4.97E-03	CCNE1	-0.72	3.76E-02	PRPS2	-0.63	1.69E-02
KRT6A	-0.93	4.25E-02	GPR19	-0.71	2.60E-02	AP1S3	-0.63	3.53E-02
SHC4	-0.93	2.88E-02	KIF5C	-0.71	1.82E-02	HNRNPD	-0.63	3.31E-03
CEACAM1	-0.92	2.67E-02	PLAGL2	-0.71	1.39E-05	SRSF12	-0.63	4.07E-02
IQCJ-SCHIP1	-0.92	1.46E-02	LIN28B	-0.71	4.84E-02	FAM92A1	-0.63	2.77E-02
TMSB15B	-0.89	8.62E-03	FZD7	-0.71	2.54E-02	FAM122B	-0.63	2.91E-02
A2ML1	-0.89	4.70E-02	SLC12A2	-0.71	4.98E-02	CEP78	-0.63	5.66E-03
PKP1	-0.87	1.77E-02	ENAH	-0.70	4.51E-03	C21orf91	-0.63	2.61E-02
GPR158	-0.87	1.65E-02	FBXO45	-0.70	1.78E-02	FRK	-0.62	3.12E-02
SLC16A1	-0.87	1.16E-02	SPC25	-0.70	2.17E-02	SLC18B1	-0.62	2.63E-02
FOLH1	-0.87	1.37E-02	ATG5	-0.70	3.36E-02	DHX36	-0.62	7.63E-04
KCNK5	-0.86	5.72E-03	DPP10-AS1	-0.70	4.60E-02	PRR16	-0.62	4.49E-02
STK39	-0.86	2.07E-03	CDH2	-0.69	4.25E-02	TTK	-0.62	3.65E-02
LOC100288911	-0.86	2.91E-02	HMGB3	-0.69	1.44E-03	PIP5K1B	-0.62	3.55E-02
OXR1	-0.85	2.77E-03	CENPV	-0.69	3.22E-02	SMIM5	-0.62	1.67E-02
CDT1	-0.85	2.33E-02	ZNF678	-0.69	1.19E-02	UXS1	-0.62	1.99E-02
BICD1	-0.85	5.88E-03	DLGAP5	-0.69	4.28E-02	NDE1	-0.62	1.06E-02
ELOVL4	-0.85	6.13E-03	SLC7A11	-0.69	4.30E-02	GMNN	-0.62	3.04E-03
E2F3	-0.85	4.40E-02	BRI3BP	-0.69	2.96E-03	CHST6	-0.61	4.76E-03

ID	log2FC	pvalue
ABCC11	1.97	1.26E-03
DHRS2	1.66	1.23E-02
REEP6	1.49	6.45E-03
AFF3	1.34	1.01E-02
SERHL	1.23	1.93E-02
HMGCS2	1.22	3.01E-02
HPD	1.20	2.62E-02
LBP	1.14	3.96E-02
AQ1	1.14	2.66E-03
CYP4Z2P	1.13	2.10E-02
THRSP	1.12	3.15E-02
ALOX15B	1.12	1.77E-02
DIO1	1.09	1.86E-02
SLC26A3	1.08	1.00E-02
UGT2B28	1.06	3.59E-02
TSPAN8	1.05	4.03E-02
GSTM5	1.04	2.68E-03
ACSM1	0.97	3.65E-02
PGC	0.94	3.36E-02
HPGD	0.94	4.82E-02
CUX2	0.91	3.10E-02
FCER1A	0.90	1.54E-02
EPYC	0.90	4.62E-02
CILP	0.90	2.69E-02
PROL1	0.88	3.66E-02
MPP7	0.88	2.10E-02
PNLIPRP3	0.88	3.00E-02
FCN2	0.87	1.26E-02
MAP7	0.86	6.97E-05
C2orf72	0.86	2.78E-03
KDM6C	0.86	1.21E-02
C4A	0.85	3.79E-02
CMBL	0.83	2.27E-02
ARX	0.83	1.07E-02
FAT3B	0.82	4.15E-04
COL6A6	0.81	2.92E-02
SUSD5	0.81	4.66E-02
SLC4A5	0.81	3.56E-02
ATP13A4	0.79	4.22E-02
GSTM1	0.79	6.56E-03
HGD	0.77	3.66E-02
TTC39A	0.77	3.02E-02
CTSO	0.76	2.79E-03
IGF2	0.76	1.49E-02
FLRT3	0.76	2.44E-02
LOC101926959	0.76	8.55E-03
GNMT	0.75	4.20E-02
MGAT2	0.73	4.70E-02
FMO5	0.73	1.58E-02
LIN7A	0.73	3.60E-03
NAT2	0.72	5.18E-03
GSTM2	0.71	1.34E-02
BRCA1	0.71	1.40E-02
NRK	0.71	2.18E-02
LINC01511	0.71	1.79E-02
TMEM51	0.71	1.50E-02
NAAA	0.71	1.63E-02
HPX	0.69	2.10E-02
LONP2	0.69	1.82E-02
SLC7A2	0.68	3.31E-02
OMD	0.68	2.58E-02
CREB3L4	0.68	1.76E-02
EFHD1	0.67	3.21E-02
PPP1R3E	0.67	9.64E-04
GPRC5C	0.67	6.40E-03
FOXO1	0.67	3.56E-02
ADAMTS2	0.67	2.59E-02
PTPN21	0.67	3.10E-02
AR	0.65	3.89E-02
TUSC5	0.64	3.65E-02
HPDL	0.64	1.28E-02
NR2F1	0.64	4.62E-02
QPRT	0.64	4.30E-02
IGSF10	0.63	8.91E-03
ACOX2	0.63	2.65E-02
LDHD	0.63	2.30E-02
PCDHA1	0.63	1.68E-02
DTNA	0.63	3.57E-02
GLYATL1	0.62	4.72E-02
MALRD1	0.62	1.24E-03

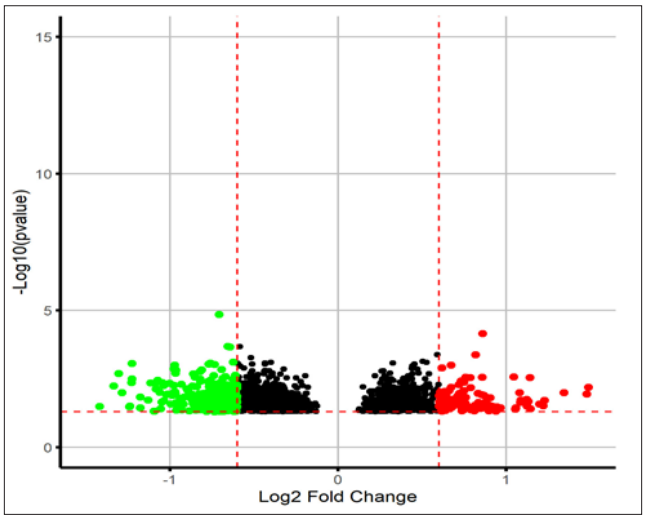


Figure 2: GPL570 samples - Genes within the range of -1.5<fold change >1.5 and p-value <0.05 were considered significantly differentially regulated. The genes with significant p-value <0.05 were extracted and shown as a volcano plot. Red colour dots denote upregulated genes and green dots denote downregulated genes in the No Response category

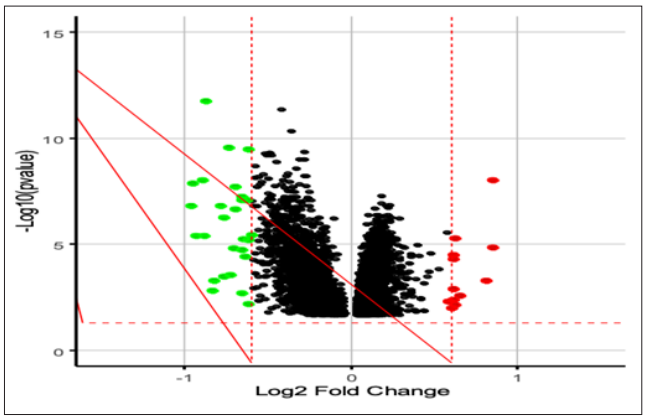


Figure 3: GPL96 samples - Genes within the range of -1.5<fold change >1.5 and p-value <0.05 were considered significantly differentially regulated. The genes with significant p-value <0.05 were extracted and shown as a volcano plot. Red colour dots denote upregulated genes and green dots denote downregulated genes in the No Response category

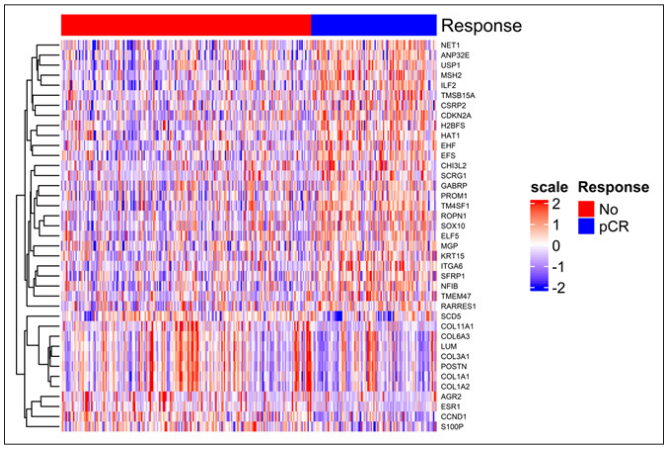


Figure 4: The Heatmap above shows the hierarchical Clustering pattern and expression of differentially expressed genes across response and no response categories of patients in GPL96 profile samples. Each column in the heatmap denotes the samples and each row denotes the expression of genes. Red colour denotes higher expression and blue colour denotes the lesser expression of genes. The red and blue lines above the heatmap gives sample information: Red -No Response Category and Blue - Pathological Complete Response Category of Patients.

Table 3: GPL96 samples List of 12 upregulated genes and 30 downregulated genes in No Response category

Genes upregulated in response category or downregulated in no response category			
		Fold change	p-value
ILF2	down	1.8306	1.76E-12
TM4SF1	down	1.9447	1.54E-07
ROPN1	down	1.9388	1.34E-08
ELF5	down	1.9095	3.94E-06
NFIB	down	1.8604	9.34E-09
SFRP1	down	1.8463	4.09E-06
GABRP	down	1.7793	0.001536
RARRES1	down	1.7686	5.14E-04
EHF	down	1.7263	1.52E-07
ANP32E	down	1.6983	5.37E-07
KRT15	down	1.6964	3.43E-04
ITGA6	down	1.6606	2.86E-10
SCRG1	down	1.6575	2.65E-04
TMSB15A	down	1.6264	1.58E-05
USP1	down	1.6200	1.96E-08
CDKN2A	down	1.6159	2.30E-07
TMEM47	down	1.5805	1.82E-05
PROM1	down	1.5795	0.00207
NET1	down	1.5780	8.20E-08
HAT1	down	1.5748	5.51E-08
CHI3L2	down	1.5602	5.75E-06
CSRP2	down	1.5523	3.77E-05
EFS	down	1.5476	8.30E-08
MSH2	down	1.5345	3.39E-10
H2BFS	down	1.5304	6.41E-06
MGP	down	1.5289	0.006553
SOX10	down	1.5175	3.56E-06
NDRG2	down	1.5088	8.64E-08
PRDX4	down	1.5059	1.55E-07
PERP	down	1.5026	5.27E-05

Genes upregulated in no response category			
		Fold change	p-value
COL11A1	up	1.5275	4.86E-05
S100P	up	1.8020	1.44E-05
CCND1	up	1.7991	9.56E-09
COL1A1	up	1.7557	5.05E-04
COL1A2	up	1.7035	9.78E-04
COL3A1	up	1.6687	3.46E-03
POSTN	up	1.5439	7.25E-03
ESR1	up	1.5414	5.28E-06
COL6A3	up	1.5348	1.24E-03
SCD5	up	1.5342	3.15E-05
AGR2	up	1.5314	4.07E-03
LUM	up	1.5182	1.03E-02

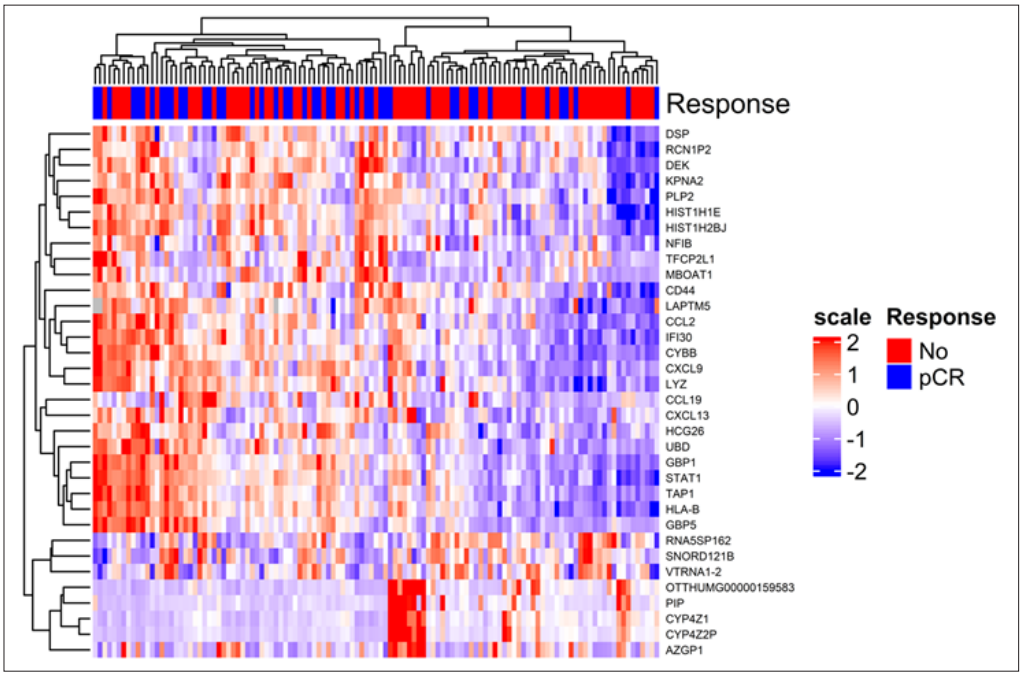


Figure 5: Heatmap of upregulated and downregulated genes in HTA2 samples

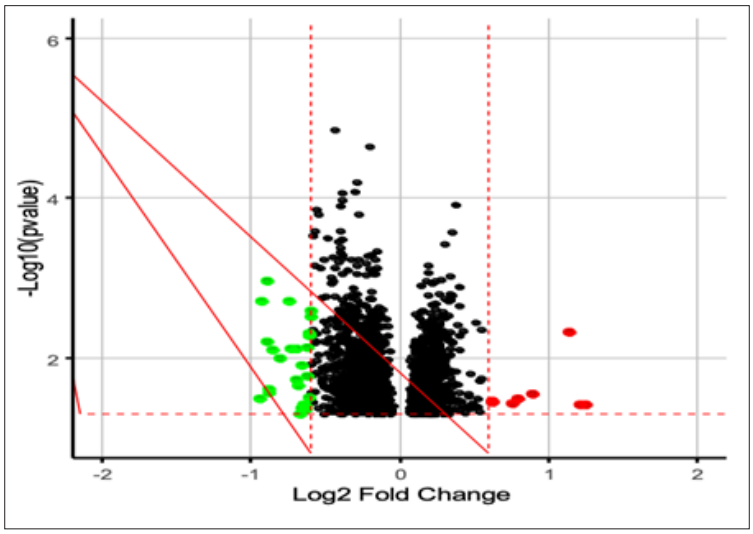


Figure 6: HTA2 samples - Genes within the range of -1.5<fold change >1.5 and p-value <0.05 were considered significantly differentially regulated. The genes with significant p-value <0.05 were extracted and shown as a volcano plot. Red colour dots denote upregulated genes and green dots denote downregulated genes in the No Response category

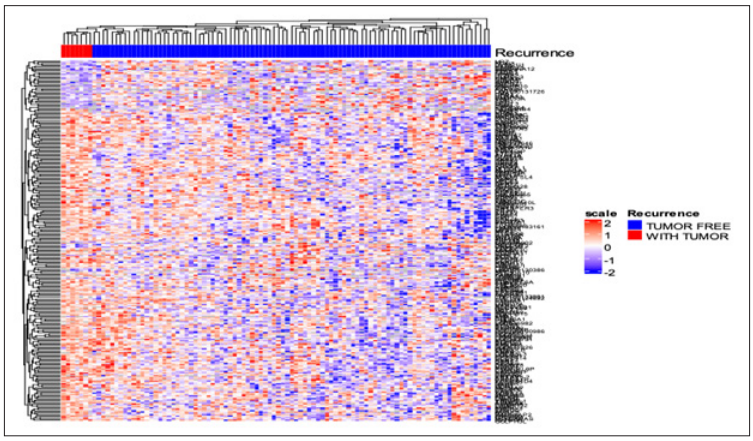


Figure 7: The heatmap above shows upregulated and downregulated genes with respect to recurrence.

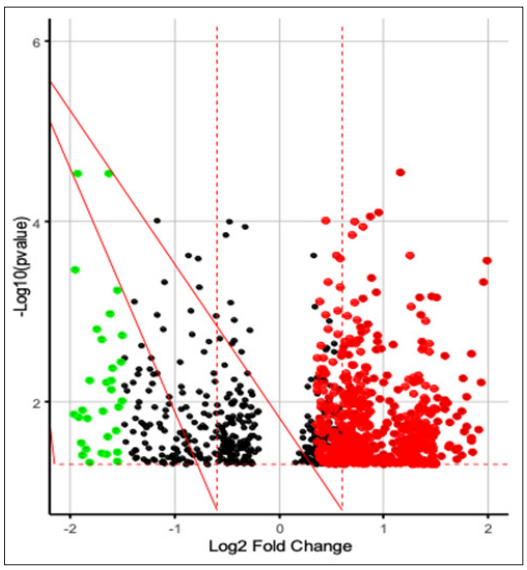


Figure 8: Volcano plot with Recurrence data - Genes within the range of -1.5<fold change >1.5 and p-value <0.05 were considered significantly differentially regulated. The genes with significant p-value <0.05 were extracted and shown as a volcano plot. Red colour dots denote upregulated genes with respect to tumours showing recurrence and green dots denote downregulated genes with respect to tumours showing recurrence.

Table 4: List of upregulated genes in Recurrence related tumors

Upregulated in Recurrence tumors (with residual tumors)

ID	log2FC	pvalue	ID	log2FC	pvalue	ID	log2FC	pvalue	ID	log2FC	pvalue
ACTL8	3.23	0.05	NR2E1	1.40	0.01	C20orf152	1.03	0.01	USP43	0.81	0.01
SLC7A4	2.96	0.02	ADAMTSL4	1.39	0.04	IQCG	1.03	0.02	KLB	0.81	0.03
PCDHA11	2.88	0.00	RHOV	1.38	0.04	MCM3APAS	1.03	0.05	ZNF626	0.81	0.03
LY6D	2.78	0.03	SFRP4	1.37	0.03	HRASLS	1.03	0.02	SLC37A1	0.81	0.00
LIN28B	2.55	0.04	DCST2	1.37	0.03	CD300LB	1.02	0.05	CELSR2	0.80	0.04
SLC34A2	2.37	0.00	KRTAP5-2	1.36	0.00	GPR157	1.01	0.02	COL10A1	0.80	0.05
GPR12	2.37	0.04	FAM157A	1.36	0.04	C15orf28	1.01	0.05	SEC16B	0.80	0.04
RSPO1	2.36	0.04	NBPF22P	1.35	0.00	ASAP3	1.00	0.01	NEAT1	0.80	0.05
ALOX12	2.34	0.05	FAM131C	1.35	0.00	C9orf150	1.00	0.04	BMS1P4	0.79	0.01
B3GALT5	2.33	0.02	WNT9A	1.35	0.01	FLJ45445	0.99	0.05	DRP2	0.79	0.00
SYT15	2.28	0.02	GUCY1B2	1.32	0.02	TTC16	0.99	0.04	CASZ1	0.78	0.00
KRT78	2.18	0.03	POM121L9P	1.30	0.02	DNHD1	0.99	0.02	ARMCX2	0.77	0.01
MSMB	2.10	0.02	CCDC114	1.29	0.01	ZYG11A	0.98	0.00	GMCL1L	0.77	0.04
BCAS1	2.06	0.00	CREB3L3	1.28	0.04	ENKUR	0.97	0.03	LOXHD1	0.77	0.01
PLAC2	2.03	0.03	TNXB	1.27	0.04	SPIN2A	0.97	0.04	CPT1C	0.77	0.04
RSPH1	1.99	0.00	TMPRSS2	1.26	0.01	FOXD4	0.96	0.01	PCDHB15	0.77	0.00
LOC220594	1.96	0.00	ATP6V1B1	1.26	0.05	ARHGEF10L	0.96	0.05	PTX4	0.75	0.02
C21orf128	1.93	0.01	RAD51AP2	1.23	0.02	C11orf63	0.95	0.00	GSPT2	0.75	0.03
FMO2	1.90	0.02	PCDHA3	1.23	0.02	MEX3A	0.95	0.00	ATL2	0.75	0.02
CRABP2	1.85	0.04	MSTN	1.23	0.02	C9orf53	0.94	0.02	TRAPPC6A	0.74	0.03
FAM106A	1.85	0.00	PCDHGB4	1.22	0.04	LOC646982	0.94	0.00	TRAM1L1	0.73	0.04
MS4A15	1.84	0.03	BCHE	1.22	0.03	LOC100190986	0.94	0.02	LOC100133991	0.73	0.03
CACNA1I	1.84	0.03	SORBS2	1.21	0.03	SNHG4	0.93	0.02	TCEA2	0.73	0.03
KCNC3	1.80	0.01	CCDC40	1.19	0.02	NAP1L5	0.93	0.00	C21orf63	0.73	0.03
GRHL3	1.80	0.02	EPS8L1	1.18	0.03	LOC283731	0.93	0.00	DMPK	0.73	0.01
LDHC	1.77	0.02	LOC650623	1.18	0.02	LGI4	0.92	0.05	TTC39B	0.72	0.01
KLHL34	1.76	0.01	C1orf228	1.17	0.03	UPK2	0.91	0.04	NBPF10	0.71	0.04
COL22A1	1.75	0.01	ZNF238	1.16	0.04	ZNF286B	0.91	0.03	C9orf117	0.71	0.03
KCNK15	1.75	0.04	TCTEX1D4	1.16	0.05	GUCY1A3	0.91	0.01	CAP2	0.71	0.00
PGPEP1L	1.72	0.03	TMEM63C	1.16	0.05	AMOT	0.90	0.03	KLF13	0.70	0.01
LRRC10B	1.70	0.04	CATSPER3	1.16	0.00	MAP3K1	0.89	0.05	PPL	0.70	0.03
DNAI1	1.70	0.01	SEMA6A	1.16	0.01	TPTE2P1	0.89	0.04	CSNK1A1L	0.70	0.03
FAM83E	1.69	0.03	C10orf68	1.15	0.03	ALDH1A3	0.89	0.05	STAT5A	0.70	0.01
CACNA1F	1.63	0.02	RTP4	1.14	0.03	TTC9	0.88	0.01	C2orf84	0.70	0.03
STRC	1.63	0.03	SLC5A6	1.14	0.01	MX2	0.88	0.00	CCDC88C	0.69	0.01
LOC100124692	1.61	0.03	ZNF883	1.13	0.02	CACNB3	0.87	0.00	RUSC1	0.69	0.01
KRT23	1.61	0.01	MACC1	1.11	0.02	ZNF135	0.86	0.02	LOC642826	0.69	0.00
PROM1	1.59	0.00	KRT80	1.11	0.03	PODN	0.86	0.01	BCL9	0.68	0.01
ZIC1	1.59	0.01	PTPRU	1.11	0.05	C1orf152	0.86	0.02	PCNT	0.67	0.01
CCDC144B	1.58	0.01	MACROD2	1.10	0.01	MLLT4	0.86	0.04	AGAP5	0.67	0.03
KIF12	1.58	0.01	LEPR	1.10	0.02	KIF16B	0.85	0.04	RNF180	0.67	0.01
DYDC2	1.55	0.04	MYCL1	1.09	0.03	ZMAT1	0.84	0.00	C1orf200	0.66	0.04
CRB2	1.52	0.01	IGFALS	1.09	0.04	FOXD4L1	0.84	0.01	MCL1	0.66	0.00
COMP	1.52	0.02	CGN	1.09	0.00	LHFPL4	0.84	0.00	IGSF9	0.66	0.01
PRB3	1.50	0.01	PVRL4	1.08	0.01	PITX1	0.83	0.00	H2BFXP	0.65	0.02
OCLM	1.50	0.00	GPR37L1	1.07	0.04	KRT8	0.83	0.04	PHLDB3	0.65	0.01
KRT19	1.46	0.00	CBLC	1.07	0.04	ZNF418	0.83	0.00	THNSL1	0.65	0.03
PRR15L	1.46	0.04	CA14	1.06	0.05	GOLGA6L5	0.82	0.02	TUBA1A	0.65	0.01
OTUD7A	1.42	0.03	LOC387646	1.05	0.01	RFPL2	0.82	0.04	MTHFR	0.64	0.05
CCND1	1.42	0.02	C9orf45	1.04	0.05	C2CD2	0.82	0.02	SPDYE7P	0.64	0.02
ZDBF2	1.42	0.02	CRIPAK	1.04	0.05	NEBL	0.81	0.03	C3orf18	0.64	0.05
MYO18A	0.63	0.01	C21orf91	0.63	0.03	ZMIZ1	0.60	0.03	FAM86B1	0.64	0.05
ASPRV1	0.63	0.03	GOLPH3L	0.62	0.01	MCM3AP	0.60	0.00	ANKRD5	0.64	0.04
ANKMY1	0.63	0.02	MAP3K9	0.62	0.00	URB1	0.60	0.00	LOC100133161	0.64	0.03
PDE4DIP	0.61	0.02	ZNF37A	0.59	0.03	TMEM86B	0.59	0.01	FNDC8	0.64	0.05
36951	0.60	0.05									

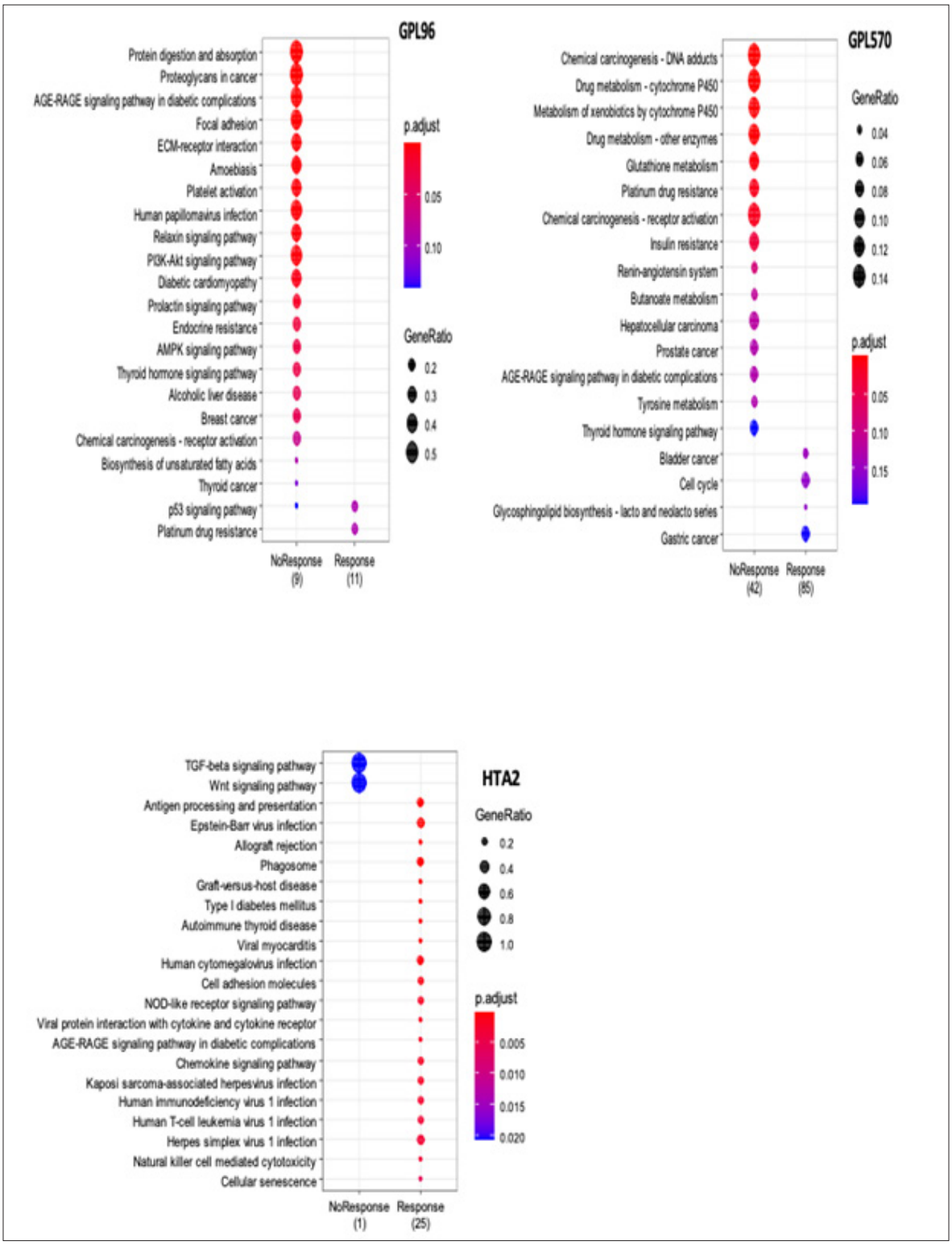


Figure 9: The gene set enrichment analysis results are shown here. It compares the gene sets enriched among upregulated and downregulated genes and plots it as Dot Plots. The size of each dot denotes the Gene ratio and the colour denotes its significance. P-adjust denotes the adjusted p value range obtained from gene set enrichment analysis.

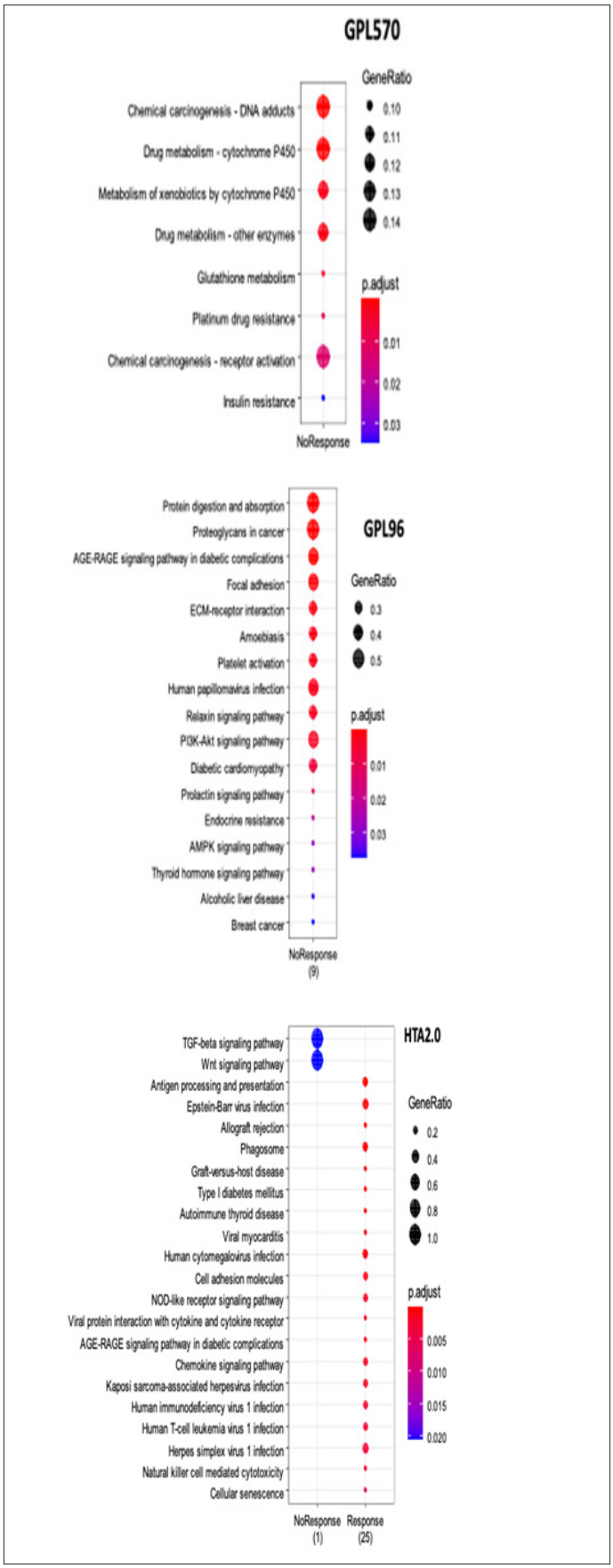


Figure 10: The plots give the significantly enriched terms. No enriched terms in response category for GPL96 and GPL570 it has not been shown.

Table 5: Important gene enrichments

Some of the important gene enrichments observed

Chemical carcinogenesis - DNA adducts	GSTM1/GSTM2/GSTM5/HPGDS/NAT2/UGT2B28
Drug metabolism - cytochrome P450	FMO5/GSTM1/GSTM2/GSTM5/HPGDS/UGT2B28
Metabolism of xenobiotics by cytochrome P450	GSTM1/GSTM2/GSTM5/HPGDS/UGT2B28
Drug metabolism - other enzymes	GSTM1/GSTM2/GSTM5/NAT2/UGT2B28
Glutathione metabolism	GSTM1/GSTM2/GSTM5/HPGDS
Platinum drug resistance	BRCA1/GSTM1/GSTM2/GSTM5
Chemical carcinogenesis - receptor activation	AR/CREB3L4/GSTM1/GSTM2/GSTM5/UGT2B28
Insulin resistance	AGT/CREB3L4/FOXO1/PPP1R3E

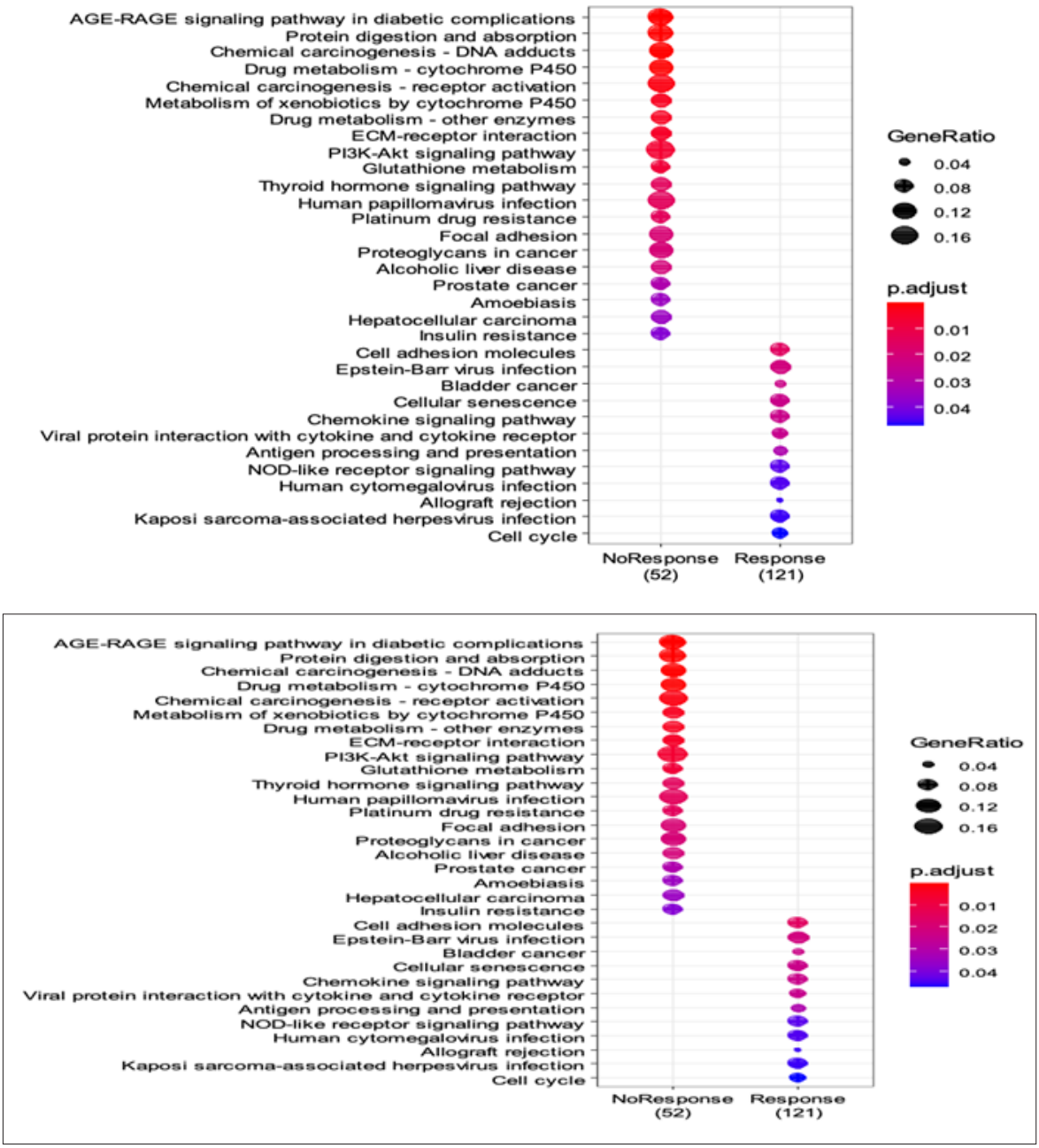


Figure 11: The pathway term enriched in the genes compiled from all three profiles. This plot shows the significant terms enriched among no response and response group of patients.

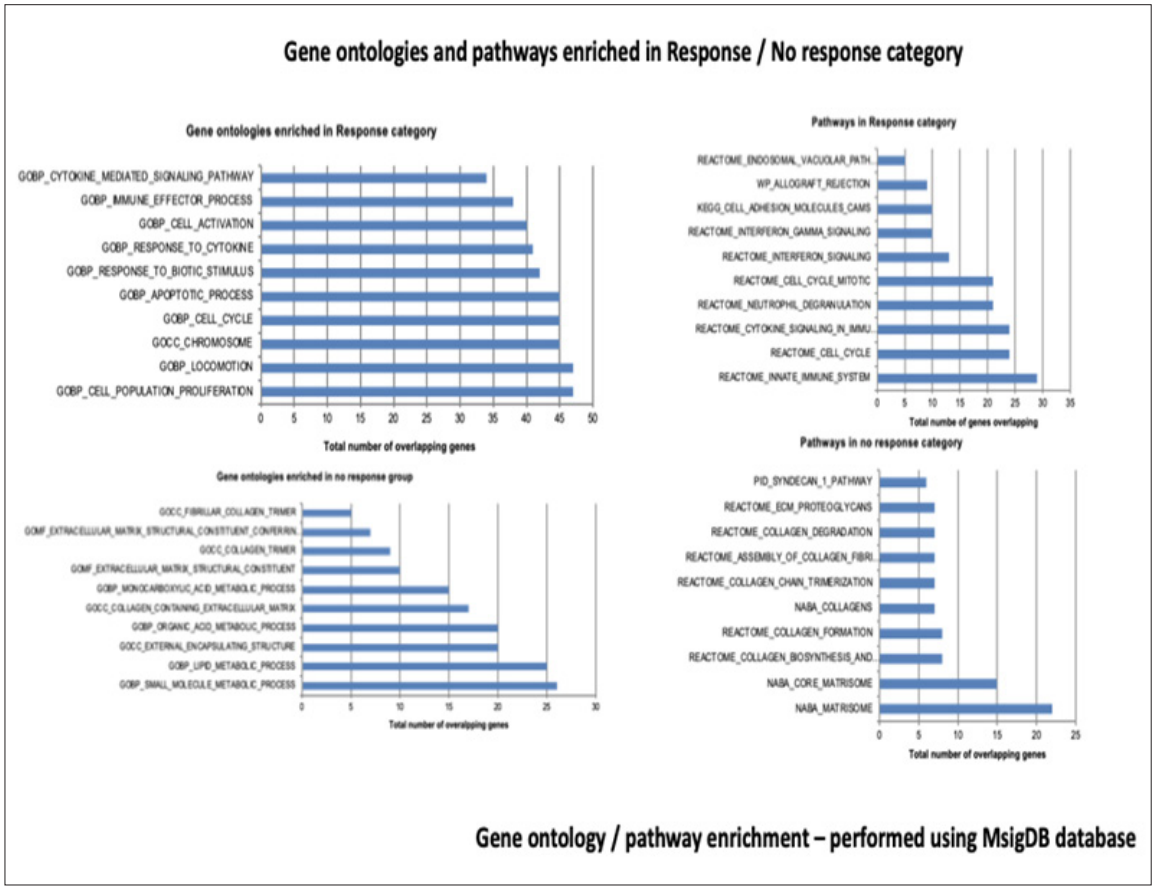


Figure 12: Gene Ontology

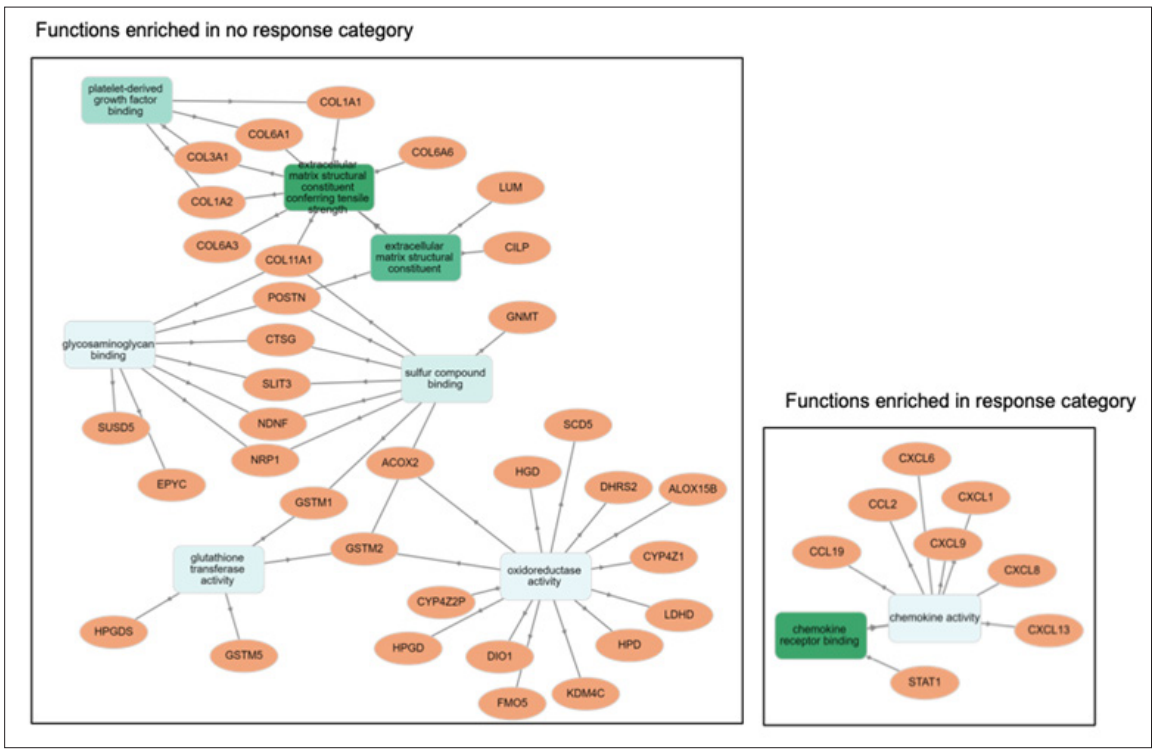


Figure 13: Functions enriched in both categories.

Discussion

Predicting individual patient responses to therapy is a major breakthrough in customizing medicine and would be of great importance for therapeutic decision-making. In this study, we identified upregulated and downregulated genes from three independent datasets with TNBC cases. High expression of the certain genes was significantly associated with a better outcome. Certain chemotherapies especially anthracyclines and taxanes were associated with higher chance of achieving pCR.

Standard therapy in cases of Triple-Negative Disease includes use of 5-Flourouracil (5-FU) Doxorubicin and Cyclophosphamide or 5-FU, Epirubicin, Cyclophosphamide or a combination of Doxorubicin and Cyclophosphamide. Paclitaxel and Docetaxel may be added to the regimen. There has been some trend towards single-agent Paclitaxel or Docetaxel as well. Literature reports the Basal Like 1 and Basal Like 2 tumour subtypes with higher expression of cell cycle and DNA damage response genes respond well to the addition of Platinum agents such as Cisplatin or Carboplatin which cause higher response rates. The Mesenchymal and Mesenchymal Like tumour subtypes respond to PI3K or mTOR inhibitor or dasatinib (abl/src inbibitor) which attack the epithelial to mesenchymal transition pathways. The Luminal Androgen Receptor Subtype with androgen receptor signaling pathway expression responds to addition of Androgen Receptor Antagonist such as bicalutamide.

Therapy guidelines has also been divided based on Genomic Profiling where the Programmed Death Ligand-1 (PDL-1) positive TNBC cases respond to addition of Immune checkpoint inhibitors in addition to standard chemotherapy. The PDL-1 negative tumours respond to single agent chemotherapy over combination (based on subtypes). In cases of prior two chemotherapies regimens having been used there is recommendation to add sacituzumab and govitecan. In cases of Germline BRCA mutations, a poly ADP-ribose polymerase inhibitor acts better over standard neoadjuvant chemotherapy.

However, despite these guidelines there is a rise in Chemotherapy resistance caused by various reasons ranging from the action of ATP Binding Cassette Proteins medicated efflux of drugs, activation of the hedgehog pathway, effect of cancer stem cells and their self-renewal, promotion of Hypoxia Inducible Factors (HIFs), hypoxia in tumour microenvironment which causes the acidic microenvironment to dampen effect of chemotherapy agents which work best in oxygen dependent manner, decreased cellular penetration of drugs, avoidance of apoptosis and dysfunction of signaling pathways. There is a need to find ways to rise up to this challenge of chemoresistance.

Literature shows, the presence of multigene signatures such as Oncotype and MammaPrint, which recommended by scientific bodies and used ny clinicians to predict prognosis of Oestrogen positive breast tumours. The measurement consists of an array of a diverse set of genes, however, their prognostic capability is largely due to their proliferation predictors. There have been instances of claims in the past where genes linked to mechanism of disease were occasionally used to predict prognosis. An example of this is the p27Kip1, which when reported as low among breast cancer cases is associated with poor survival [17]. We intended to obtain a similar prognostic capacity by analysing the gene expression profiles in cases of Triple- Negative Breast disease.

As explained in the Results section above, there was little overlap in genes among the different datasets even though the predictive power is similar. Previously too, concern has been voiced regarding the uniqueness of signature genes in predicting outcomes in TNBC [18]. To date, no multigene assay are recognized or ratified by clinical guidance for predicting the prognosis of triple-negative breast cancers [19]. In our study, we found in the Platinum-based cohort the enrichment of key terms like Platinum drug resistance, chemical carcinogenesis, glutathione metabolism-related terms in no response category. The Taxane-based cohort revealed a defined enrichment of oncogenic signaling pathways like Pi3K-AKT, ECM, Focal adhesion etc., in the non response category and the Anthracycline±taxane cohort results were not highly significant as the total number of differentially regulated genes were very less, it provide us an idea on the action of immune response pathways in conferring better immune response to TNBC patients.

In the GPL 570 category results, the most salient genes upregulated in the no response category were ABCC11/MRP8 which confers resistance to anthracyclines, taxanes, mitoxantrone, 5 Fluorouracil through the Multidrug resistance gene Protein 8, FCER1A seen in the relapse group in previous research, GSTM5 which is associated with brain metastasis, LBP associated with Pregnancy-associated TNBC, TSPAN8 associated with shorter overall survival, EDNRB - High expression levels of which in response category were associated with favourable disease-free survival time and PCDHB2 which is a known Prognostic marker for aggressive Basal-like phenotype. Each of these can serve as novel prognostic biomarkers.

Under the GPL96 results the strongest association in the no response category was shown by COLL11A1. The ECM proteome, or "matrisome," of cancers has been an area of extensive research recently. Matrisome is defined as a collection of genes encoding core ECM proteins and ECM-associated proteins. Chemotherapy and immunotherapy resistance associated with COL11A1 expression is increased in several cancers and high levels of COL11A1 are often associated with poor survival, chemoresistance, and recurrence. COL11A1 also promotes cancer cell migration, metastasis, and therapy resistance by activating pro-survival pathways and modulating tumor metabolic phenotype. Several inhibitors that are currently being tested in clinical trials for cancer or used in clinic for other diseases, can be potentially used to target COL11A1 signaling. COL11A1 as a promising biomarker and a key player in cancer.

In the HTA2 array dataset downregulation of proteins in the major histocompatibility complex (MHC) class I antigen presentation pathways were enriched, including TAP1 and were associated with significantly shorter recurrence-free and overall survival. TNBC cells associated with HLA-B protein express at least one predicted neoepitope, indicating that most breast cancer patients have the potential for neo-epitope targeted immunotherapy. Neoepitopes are targets for vaccine development and adoptive T cell therapy. The combination of checkpoint blockade with a patient-specific neoepitope vaccines are being developed. DEK promoted cancer cell angiogenesis and metastasis by activating the PI3K/AKT/mTOR pathway and therefore, it might be a potential target in TNBC therapy.

Amongst the pathway enriched for significant terms the most interesting find was RAGE signaling pathway and could be used as a novel therapeutic target and biomarker for TNBCs. The RAGE/S100A7 signaling acts through paracrine and endocrine mechanisms in epithelial as well as Tumour associated macrophages and plays an important role in linking inflammation

to breast cancer metastasis. In addition, RAGE also binds to S100A8/A9 to recruit myeloid-derived suppressor cells (MDSCs) and thereby enhance breast cancer growth and metastasis further.

The most important gene enrichments in the No response category were Drug Resistance and Glutathione Metabolism related pathways. Drug Resistance especially Cisplatin resistance has been elucidated through cell death inhibition (apoptosis suppression), altering in the drug metabolism, epigenetic changing, changes in the drug targets, enhanced DNA-repair and target gene amplification. Drug resistance is seen with the use of Platinum-based therapy such as Paclitaxel which enhanced the expression of VEGF receptor 1 expressed on lung endothelial cells causing lung metastasis. Also, Doxorubicin which enhances the perivascular TIE-2 expression on macrophages causing extravasation and proliferation of tumour in extracellular matrix. Chemotherapeutic drugs can cause pro-apoptotic exosomal extracellular vesicle generation as well. Glutathione biosynthesis is caused due to Integrated stress response which activates a signaling pathway. The phosphorylation of eIF2-alpha in TNBC cells promotes activation of ATF4 with reduced glutathione within cells causing more Reactive oxygen Species associated damage.

In the recurrence category due to the small sample size, we could not use it to draw meaningful inferences. However, some validated by studies show ACTL8, which was identified to be associated with recurrence in these samples, may facilitate the proliferation, migration and invasion, while inhibiting apoptosis through activating PI3K/Akt/mTOR signaling pathway in TNBC. ACTL8 might be considered a novel prognostic marker and therapeutic target for TNBC.

The solute carrier (SLC) 7 family genes play central roles in cancer cell metabolism as glucose and glutamate transporters. SLC7A had excellent diagnostic value in BC patients and was strongly correlated with tumour infiltration. SLC7A4 expression levels can be sensitive biomarkers for predicting BC outcomes.

Selection of optimal chemotherapy for a patient is a challenging issue. One of the many aims by the scientific community in improving prediction and response behaviour, has been focusing on pCR responses, whether their better responses could be achieved by predicting response to specific chemotherapies. Individual patients respond differently to various therapy cycles and treatment regimens; for example, addition of taxanes into anthracycline regimens is known to improve the efficacy of chemotherapy and even among subgroups in triple-negative breast cancer, a more complete pCR rate could be achieved when treated by certain regimens [20-25].

Essential work done by Shawn McGuirk et al and Sherene Loi et al. have shown how chemotherapy resistance is a critical therapeutic barrier [26,27]. Metabolic adaptations have been the source of resistance with specific therapies eliciting unique metabolic alterations. The dependence of the breast cancer cells on distinct metabolic adaptations reveals that these can vary with different therapeutic modalities and that these metabolic dependencies can be exploited as a targeted approach to treat chemotherapy-resistant breast cancer.

Limitations of the study include that the datasets are not representative of the whole picture of TNBC patients. The datasets are unwieldy to work with and need a lot of cleaning and data management resources. Datasets carry limited variables of interest. The data may be subject to clerical error and coding inaccuracy. There may be the omission of useful data due to perceived irrelevance. The lack of follow-up data limits our ability to make associations between genetic alterations and clinical outcomes. The use of therapy-specific data may act as a potential confounder. Available information is of limited granularity and de-identification precludes the collection of additional variables thus increasing the risk of unmeasured confounding. Follow-up data limits the ability to make clinically meaningful associations between genetic alterations and clinical outcomes. The selection of different treatment regimens can act as a potential confounder. The differentiation of causation genes versus correlation genes can be difficult. Loss of information due to limiting network to most dense cores.

Recurrence data was too small which was a major limitation. Attempts to improve cancer recurrence information should focus on hospital operational and process factors surrounding how the hospital tumour registries collect recurrence data and make it available to the research community for analysis.

However, the study helps us to better identify the actionable mutations and genes related to response to therapy which can have immense use in matching therapy to each individual patient and better overall survival in TNBC patients. Predictive biomarkers of drug therapy are essential to the treatment of TNBC patients. The distinct difference in the genetic mutations between responding and non-responding patients proves the usefulness of this process to predict Chemotherapy efficiency. The somatic mutations post-chemotherapy were not well elucidated previously in any other study, to the best of our knowledge. No response genes can highlight highly deregulated pathways early which can act as markers. Any upregulated genes can be studied early using biopsy of TNBC tissues. Alternate therapy for genes or pathways can be targeted. Cross-Talk between signaling pathways paves way to experiment more combination therapies. Our study highlights some of these and helps in contributing to the knowledge required for guiding therapy and predicting responses to given therapies by understanding the genetically variable characteristics. Commonly seen as an umbrella term for a heterogeneous disease, TNBC, genetic profiling shall allow us to sequence individual genomes of patients which shall usher us into the era of personalized treatment guided by genetic analysis of each patient's individual tumours as key for better therapy selection.

Conclusion and Future Directions

Our study demonstrates the successful identification of responder and no-response genes to predict treatment response to specific chemotherapy in Triple-negative breast Disease. The dependence of the breast cancer cells on distinct metabolic adaptations reveals that these can vary with different therapeutic modalities and that these metabolic dependencies can be exploited as a targeted approach to treat chemotherapy-resistant breast cancer.

Data Availability

All raw are available through NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) as detailed in Supplemental Table 1.

Contributions

A.B. and D.S conceived and designed the study; D.S collected and analysed data; A.B. and D.S. interpreted the results; D.S. wrote the manuscript, and A.B and D.S revised the manuscript; Both the authors reviewed and approved the manuscript.

Ethical Interests

The authors have no competing interests

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, et al. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 71: 209-249.
2. Gluz O, Liedtke C, Gottschalk N, Pusztai L, Nitz U, et al. (2009) Triple-negative breast cancer--current status and future directions. Ann Oncol 20:1913-1927.
3. Foulkes WD, Smith IE, Reis-Filho JS (2010) Triple-negative breast cancer. N Engl J Med 363: 1938-1948.
4. Jiang YZ, Ma D, Suo C, Shi J, Xue M, et al. (2019) Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell 35: 428-440.
5. Bareche Y, Venet D, Ignatiadis M, Aftimos P, Piccart M, et al. (2018) Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. Ann Oncol 29: 895-902.
6. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy Ab, et al. (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121: 2750-2767.
7. Pareja F, Reis-Filho JS (2018) Triple-negative breast cancers - a panoply of cancer types. Nat Rev Clin Oncol 15: 347-348.
8. Turner NC, Reis-Filho JS (2013) Tackling the diversity of triple-negative breast cancer. Clin Cancer Res 19: 6380-6388.
9. Staaf J, Glodzik D, Bosch A, Vallon-Christersson J, Reutersward C, et al. (2019) Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat Med 25: 1526-1533.
10. Early Breast Cancer Trialists' Collaborative G (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365: 1687-1717.
11. Kozick Z, Hashmi A, Dove J, Hunsinger M, Arora T, et al. (2017) Disparities in compliance with the Oncotype DX breast cancer test in the United States: A National Cancer Data Base assessment. Am J Surg 215: 686-692.
12. Denduluri N, Somerfield MR, Chavez-MacGregor M, Comander AH, Dayao Z, et al. (2016) Selection of optimal adjuvant chemotherapy regimens for human epidermal growth factor receptor 2 (HER2)-negative and adjuvant targeted therapy for HER2-positive breast cancers. An American Society of Clinical Oncology guideline adaptation of the Cancer Care Ontario clinical practice guideline J Clin Oncol 39: 685-693.
13. Sicklick JK, Kato S, Okamura R, Schwaederle M, Hahn ME, et al. (2019) Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. Nat Med 25: 744-750.
14. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, et al. (2016) Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol 34: 1134-1150.
15. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, et al. (2008) Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 10: R65.
16. Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, et al. (2010) Genomic index of sensitivity to endocrine therapy for breast cancer. J Clin Oncol 28: 4111-4119.
17. Porter PL, Malone KE, Heagerty PJ, Alexander GM, Gatti LA, et al. (1997) Expression of cell-cycle regulators p27 Kip1 and cyclin E, alone and in combination, correlate with survival in young breast cancer patients. Nat Med 3: 222-225.
18. Ein-Dor L, Kela I, Getz G, Givol D, Domany E (2004) Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 21: 171-178.
19. Ramsey IC, Stuart AM, Shelley P, Amit G, James H, et al. (2018) Opportunities and priorities for breast surgical research. Lancet Oncol 19: 521-533.
20. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365: 1687-1717.
21. Von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, et al. (2012) Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 30: 1796-1804.
22. Asselain B, John B, Jonas B, Elizabeth BN, Judith B, et al. (2018) Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomised trials. Lancet Oncol 19: 27-39.
23. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, et al. (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 384: 164-172.
24. Gianni L, Baselga J, Eiermann W, Porta VG, Semiglazov V, et al. (2009) Phase III trial evaluating the addition of paclitaxel to doxorubicin followed by cyclophosphamide, methotrexate, and fluorouracil, as adjuvant or primary systemic therapy: European Cooperative Trial in Operable Breast Cancer. J Clin Oncol 27: 2474-2481.
25. von Minckwitz G, Schneeweiss A, Loibl S, Salat C, Denkert C, et al. (2014) Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): a randomised phase 2 trial. Lancet Oncol 15: 747-756.
26. McGuirk S, Audet-Delage Y, Annis MG, Xue Y, Vernier M, et al. (2021). Resistance to different anthracycline chemotherapeutics elicits distinct and actionable primary metabolic dependencies in breast cancer. ELife 10:e65150.
27. Loi S, Pommey S, Haibe-Kains B, Beavis PA, Darcy PK, et al. (2013) CD73 promotes anthracycline resistance and poor prognosis in triple negative breast cancer. Proceedings of the National Academy of Sciences 110: 11091-11096.