

三阴乳腺癌的多药耐药的靶点分析

2024-05-21

LiChuang Huang



@ 立效研究院

Contents

1 摘要	1
1.1 生信需求	1
1.2 结果	1
1.3 其他要求	1
1.4 其他要求的结果	1
1.5 补充分析	1
1.6 补充分析结果	1
2 前言	1
3 材料和方法	1
3.1 材料	1
3.2 方法	1
4 分析结果	2
5 结论	2
6 附：分析流程	2
6.1 三阴乳腺癌	2
6.2 多药耐药	3
6.3 交集基因的富集分析	4
6.4 三个所选基因的联系	7
6.4.1 StringDB	7
7 附：分析流程	9
7.1 TCGA-BRCA	9
7.1.1 TNBC	9
7.1.2 TNBC 紫杉醇耐药性分析	10
7.1.3 差异分析	12
Reference	15

List of Figures

1	Intersection of MDR with TNBC	4
2	KEGG enrichment	5
3	GO enrichment	6
4	Hsa05206 visualization	7
5	Selected genes Top10 interaction	8
6	QQ plot for distribution of the transformed IC50 data	10
7	BR estimate prediction accuracy	11

8	BR Resistance vs Non resistance DEGs	13
---	--	----

List of Tables

1	TNBC related targets from GeneCards	2
2	MDR related targets from GeneCards	3
3	Selected genes Top20 interaction data	8
4	TNBC annotation	9
5	BR predicted drug sensitivity	11
6	Metadata	12
7	BR data Resistance vs Non resistance DEGs	14
8	BR data Resistance vs Non resistance DEGs ABCB1	15

1 摘要

1.1 生信需求

三阴乳腺癌的多药耐药的靶点分析 (创新性比较好的通路)

1.2 结果

经查阅资料,发现 MDR 所能应用的数据库或方法比较有限,难以拓展分析。以下采用了比较简单的办法得出结果,仅供参考。

- 分别对 MDR 和 TNBC 使用 GeneCards 获取相关基因,见 Tab. 2 和 Tab. 1
- 取交集基因 Fig. 1
- 对交集基因做富集分析见 Fig. 2 和 Fig. 3。
- “MicroRNAs in cancer”可能是良好的候选通路,见 Fig. 4 中的 “breast cancer” 部分。

1.3 其他要求

在对 MDR 和 TNBC 基因预测并且取交集获得靶点基因的基础上,需要找到本课题所研究的 ABCB1/YBX1/BCL2 轴即关注 ABCB1 和 YBX1 基因的下游信号通路,通过 GO 富集分析以及 KEGG 富集分析预测 ABCB1/YBX1 和 BCL2 之间的关联

1.4 其他要求的结果

见 6.4。

1.5 补充分析

使用临床数据,通过对三阴乳腺癌和癌旁组织进行生信分析,找到其中的关于紫杉类药物耐药的差异基因 ABCB1 (此为需要的目的基因)

1.6 补充分析结果

成功筛选到 ABCB1,见 Tab. 8。

其余信息见 7

2 前言

3 材料和方法

3.1 材料

3.2 方法

Mainly used method:

- R package `ClusterProfiler` used for gene enrichment analysis¹.
- The Human Gene Database `GeneCards` used for disease related genes prediction².
- R Package `pRRophetic` was used for Prediction of Clinical Chemotherapeutic Response³.
- R package `STEINGdb` used for PPI network construction^{4,5}.
- R package `pathview` used for KEGG pathways visualization⁶.
- The MCC score was calculated referring to algorithm of `CytoHubba`⁵.
- R version 4.4.0 (2024-04-24); Other R packages (eg., `dplyr` and `ggplot2`) used for statistic analysis or data visualization.

4 分析结果

5 结论

6 附：分析流程

6.1 三阴乳腺癌

Table 1 (下方表格) 为表格 TNBC related targets from GeneCards 概览。

(对应文件为 `Figure+Table/TNBC-related-targets-from-GeneCards.xlsx`)

注：表格共有 491 行 7 列，以下预览的表格可能省略部分数据；含有 491 个唯一 ‘Symbol’。

The GeneCards data was obtained by querying :

Triple negative breast cancer

Restrict (with quotes) :

TRUE

Filtering by Score: :

Score > 3

Table 1: TNBC related targets from GeneCards

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
BRCA1	BRCA1 DNA ...	Protein Co...	P38398	59	GC17M043044	29.76
BARD1	BRCA1 Asso...	Protein Co...	Q99728	55	GC02M214725	19.27
BRCA2	BRCA2 DNA ...	Protein Co...	P51587	56	GC13P032315	19.14
EGFR	Epidermal ...	Protein Co...	P00533	63	GC07P055019	17.03

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
TP53	Tumor Prot...	Protein Co...	P04637	62	GC17M007661	15.21
CD274	CD274 Mole...	Protein Co...	Q9NZQ7	54	GC09P005450	14.49
PALB2	Partner An...	Protein Co...	Q86YC2	53	GC16M023603	13.77
LOC126862571	BRD4-Indep...	Functional...		9	GC17P103838	13.42
LINC01672	Long Inter...	RNA Gene		18	GC01P011469	11.84
CHEK2	Checkpoint...	Protein Co...	O96017	63	GC22M028687	11.81
AR	Androgen R...	Protein Co...	P10275	60	GC0XP067544	11.11
H19	H19 Imprin...	RNA Gene		34	GC11M001995	11.05
LDHA	Lactate De...	Protein Co...	P00338	58	GC11P018394	10.71
ERBB2	Erb-B2 Rec...	Protein Co...	P04626	63	GC17P039687	10.66
STAT3	Signal Tra...	Protein Co...	P40763	62	GC17M042313	10.6
...

6.2 多药耐药

Table 2 (下方表格) 为表格 MDR related targets from GeneCards 概览。

(对应文件为 Figure+Table/MDR-related-targets-from-GeneCards.xlsx)

注：表格共有 722 行 7 列，以下预览的表格可能省略部分数据；含有 722 个唯一 ‘Symbol’。

The GeneCards data was obtained by querying :

Multidrug Resistance

Restrict (with quotes) :

TRUE

Filtering by Score: :

Score > 1

Table 2: MDR related targets from GeneCards

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
ABCB1	ATP Bindin...	Protein Co...	P08183	60	GC07M087504	66.16
ABCC1	ATP Bindin...	Protein Co...	P33527	56	GC16P015949	63.99
ABCC2	ATP Bindin...	Protein Co...	Q92887	57	GC10P099782	47.35
ABCG2	ATP Bindin...	Protein Co...	Q9UNQ0	58	GC04M088090	30.63
ABCC3	ATP Bindin...	Protein Co...	O15438	53	GC17P050634	29.32

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
ABCC4	ATP Bindin...	Protein Co...	O15439	53	GC13M095019	27.78
ABCB4	ATP Bindin...	Protein Co...	P21439	55	GC07M087365	27.09
MVP	Major Vaul...	Protein Co...	Q14764	49	GC16P065989	23.3
ABCC5	ATP Bindin...	Protein Co...	O15440	52	GC03M183919	22.16
ABCB11	ATP Bindin...	Protein Co...	O95342	55	GC02M168922	21.17
ABCC6	ATP Bindin...	Protein Co...	O95255	56	GC16M018124	18.44
ABCC10	ATP Bindin...	Protein Co...	Q5T3U5	42	GC06P043427	16.93
C19orf48P	Chromosome...	Pseudogene		30	GC19M050797	14.79
DNAH8	Dynein Axo...	Protein Co...	Q96JB1	47	GC06P125656	11.7
RPSA	Ribosomal ...	Protein Co...	P08865	55	GC03P039406	10.85
...

6.3 交集基因的富集分析

Figure 1 (下方图) 为图 Intersection of MDR with TNBC 概览。

(对应文件为 [Figure+Table/Intersection-of-MDR-with-TNBC.pdf](#))

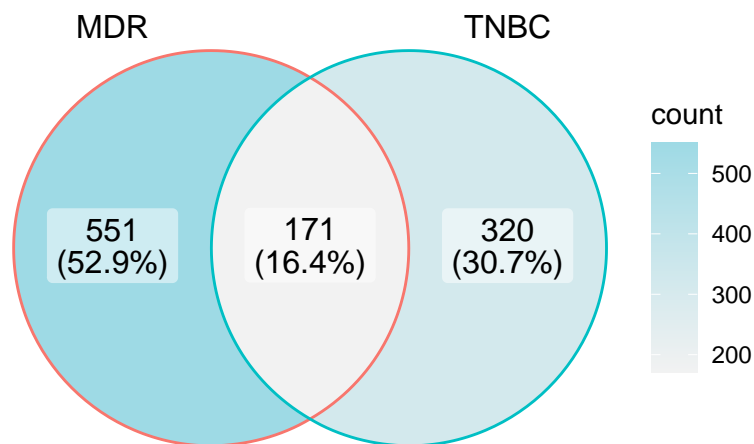


Figure 1: Intersection of MDR with TNBC

Intersection :

ABCB1, GSTP1, YBX1, LINC01672, BCL2, TP53, TOP2A, TMX2-CTNND1, ESR1, HIF1A, SCARNA5, PTGS2, AKT1, BIRC5, PVT1, CERN3, MIR7-3HG, JUN, CD44, STAT3, MIR381, PTEN, TNF, S100A4, MGMT, CAV1, MYC, EGFR, ERCC1, H19, SIRT1, SOD2-OT1, NFKB1, IL6, HSPA4, PARP1, NOTCH1, CTNNB1, VEGFA, CDH1, VIM, ANXA5, ALDH...

(上述信息框内容已保存至 [Figure+Table/Intersection-of-MDR-with-TNBC-content](#))

Figure 2 (下方图) 为图 KEGG enrichment 概览。

(对应文件为 `Figure+Table/KEGG-enrichment.pdf`)

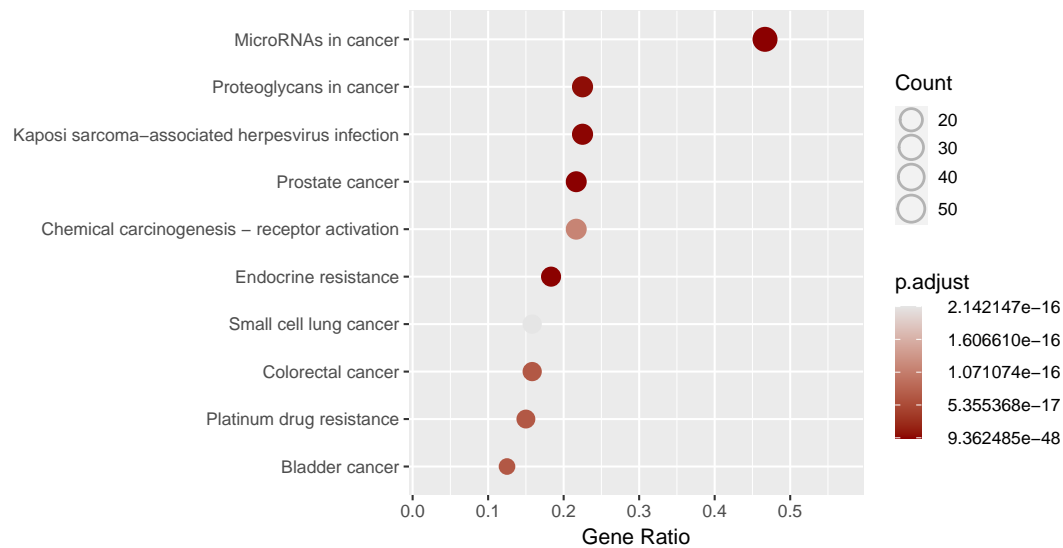


Figure 2: KEGG enrichment

Figure 3 (下方图) 为图 GO enrichment 概览。

(对应文件为 `Figure+Table/GO-enrichment.pdf`)

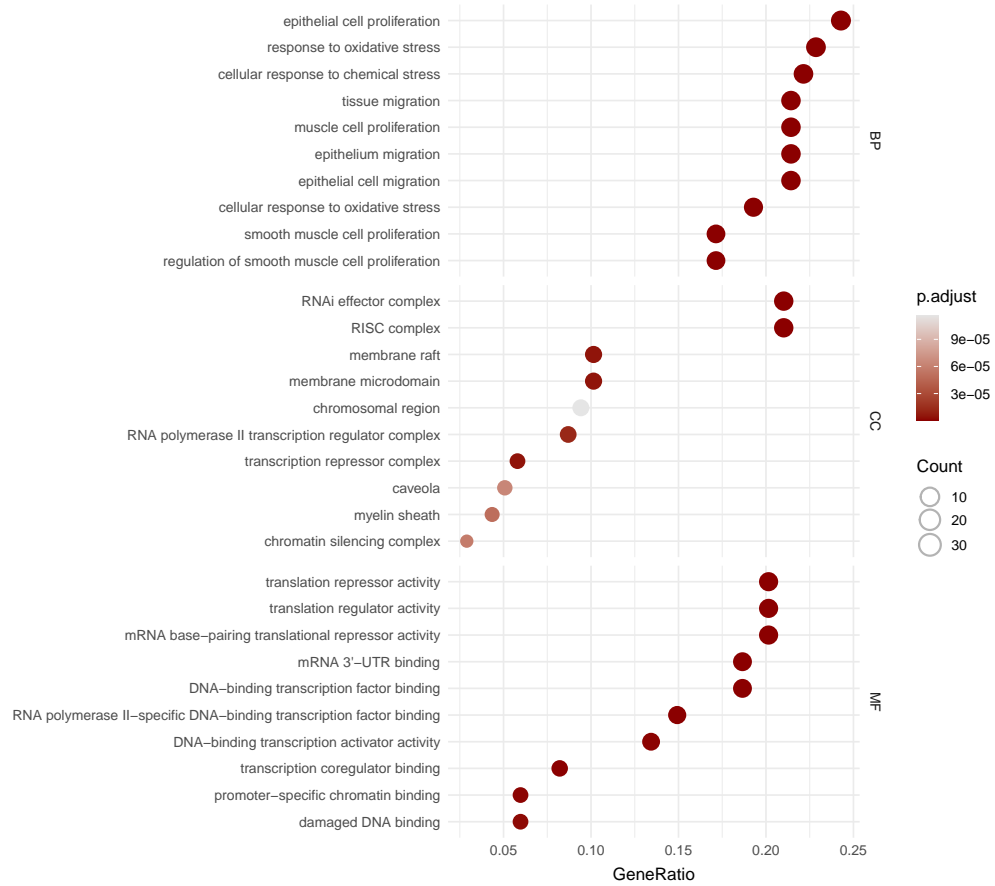


Figure 3: GO enrichment

Figure 4 (下方图) 为图 Hsa05206 visualization 概览。

(对应文件为 Figure+Table/Hsa05206-visualization.png)

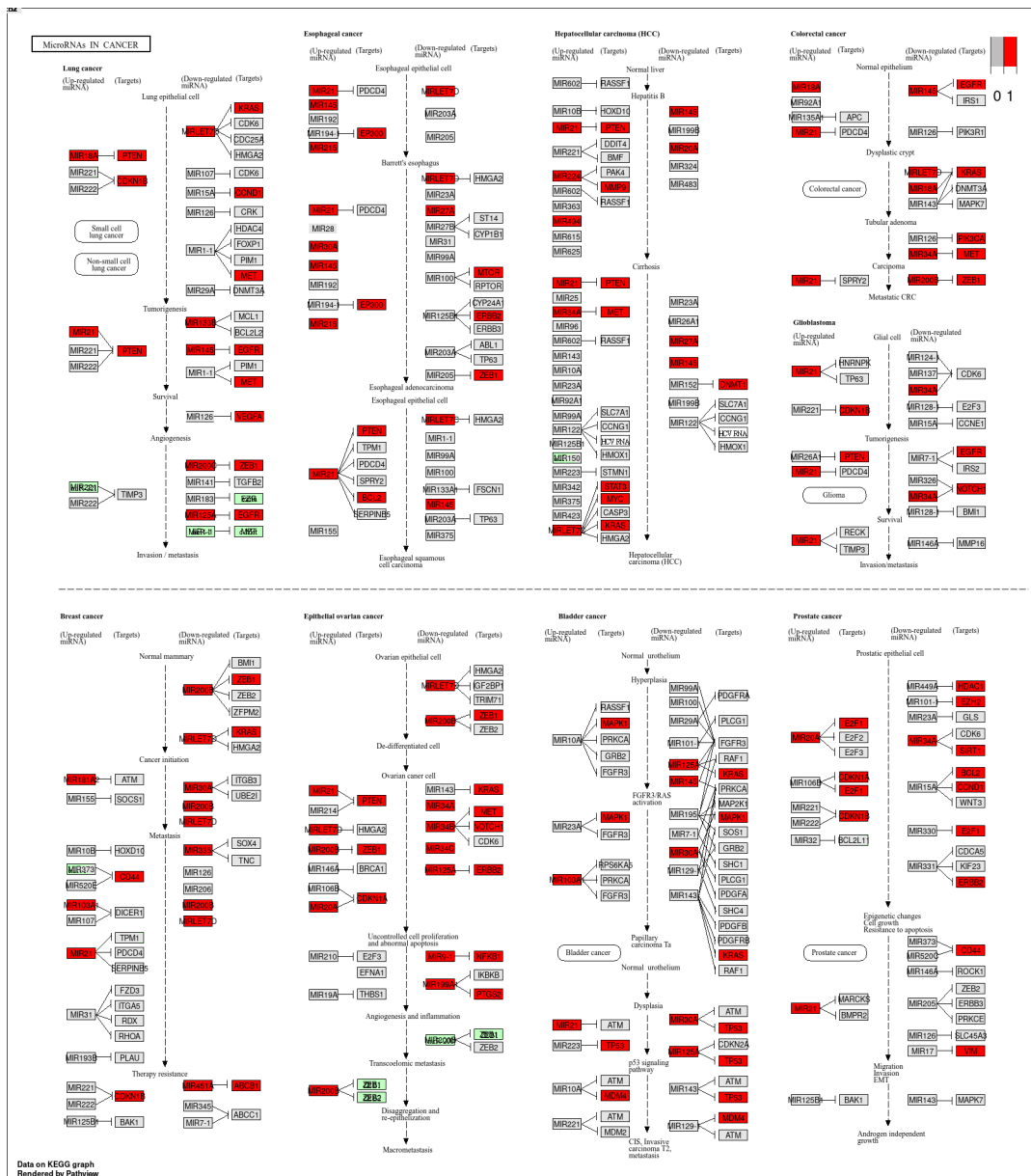


Figure 4: Hsa05206 visualization

Interactive figure :

<https://www.genome.jp/pathway/hsa05206>

6.4 三个所选基因的联系

6.4.1 StringDB

以 STRINGdb 对 Fig. 1 构建 PPI 网络 (physical, 可直接相互作用的网络), 获取 MCC top 10 的蛋白, 重新构建这些蛋白和 ABCB1, YBX1, BCL2 的 PPI 网络, 见 Fig. ??。

Figure 5 (下方图) 为图 Selected genes Top10 interaction 概览。
 (对应文件为 `Figure+Table/Selected-genes-Top10-interaction.pdf`)

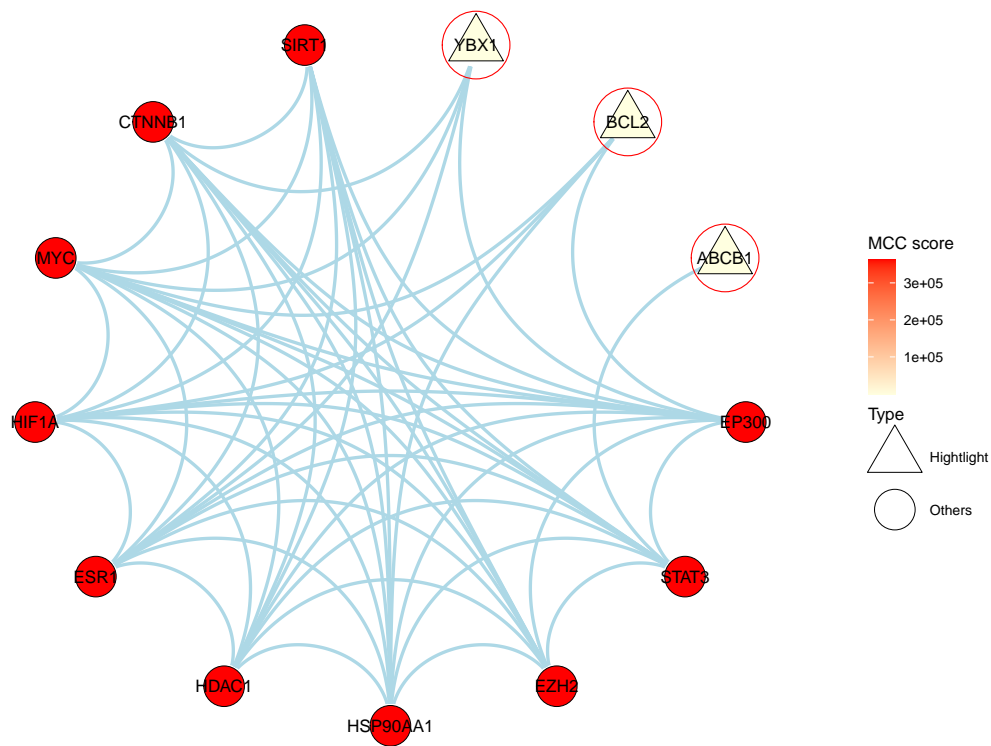


Figure 5: Selected genes Top10 interaction

Table 3 (下方表格) 为表格 Selected genes Top20 interaction data 概览。
 (对应文件为 `Figure+Table/Selected-genes-Top20-interaction-data.csv`)

注：表格共有 54 行 2 列，以下预览的表格可能省略部分数据；含有 12 个唯一 ‘Source’。

Table 3: Selected genes Top20 interaction data

Source	Target
EP300	SIRT1
STAT3	SIRT1
STAT3	EP300
EZH2	SIRT1
EZH2	EP300
EZH2	STAT3
HSP90AA1	SIRT1

Source	Target
HSP90AA1	EP300
HSP90AA1	STAT3
HSP90AA1	EZH2
YBX1	EP300
HDAC1	SIRT1
HDAC1	EP300
HDAC1	STAT3
HDAC1	EZH2
...	...

7 附：分析流程

7.1 TCGA-BRCA

数据来源于 TCGA-BRCA

7.1.1 TNBC

获取 TCGA-BRCA 的标释，取 TNBC 子集。

Table 4 (下方表格) 为表格 TNBC annotation 概览。

(对应文件为 **Figure+Table/TNBC-annotation.xlsx**)

注：表格共有 1059 行 45 列，以下预览的表格可能省略部分数据；含有 1059 个唯一 ‘TCGA_SAMPLE’。

Table 4: TNBC annotation

TCGA_S...	BARCODE	TNBC	PAM50	PAM50lite	TNBCtype	TNBCty...	IM_cen...	MSL_ce...	BL1_ce...
TCGA-A...	TCGA-A...	YES	Basal	Basal	UNC	BL1	-0.067...	-0.204...	0.0901...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	0.0890...	-0.409...	0.6770...
TCGA-A...	TCGA-A...	YES	Basal	Basal	IM	BL1	0.5766...	-0.304...	0.3889...
TCGA-A...	TCGA-A...	YES	Basal	Basal	UNC	BL1	0.0583...	-0.299...	0.1577...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	-0.036...	-0.184...	0.2627...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	0.1164...	-0.415...	0.5891...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	-0.282...	-0.016...	0.4580...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	0.3572...	-0.209...	0.4447...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	-0.273...	-0.280...	0.6290...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	0.0485...	-0.160...	0.2680...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	-0.099...	-0.368...	0.5270...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	-0.299...	-0.358...	0.5655...

TCGA_S...	BARCODE	TNBC	PAM50	PAM50lite	TNBCtype	TNBCty...	IM_cen...	MSL_ce...	BL1_ce...
TCGA-A...	TCGA-A...	YES	Basal	Basal	BL1	BL1	0.1769526	-0.267...	0.6534...
TCGA-A...	TCGA-A...	YES	Basal	Basal	IM	BL1	0.6529...	-0.285...	0.4433...
TCGA-A...	TCGA-A...	YES	Basal	Basal	MSL	BL1	0.1206...	0.2856...	0.0313...
...

7.1.2 TNBC 紫杉醇耐药性分析

使用 pRRophetic 预测紫杉醇 Paclitaxel 耐药性 (IC50)，并根据 IC50 分值分组。

Figure 6 (下方图) 为图 QQ plot for distribution of the transformed IC50 data 概览。

(对应文件为 [Figure+Table/QQ-plot-for-distribution-of-the-transformed-IC50-data.pdf](#))

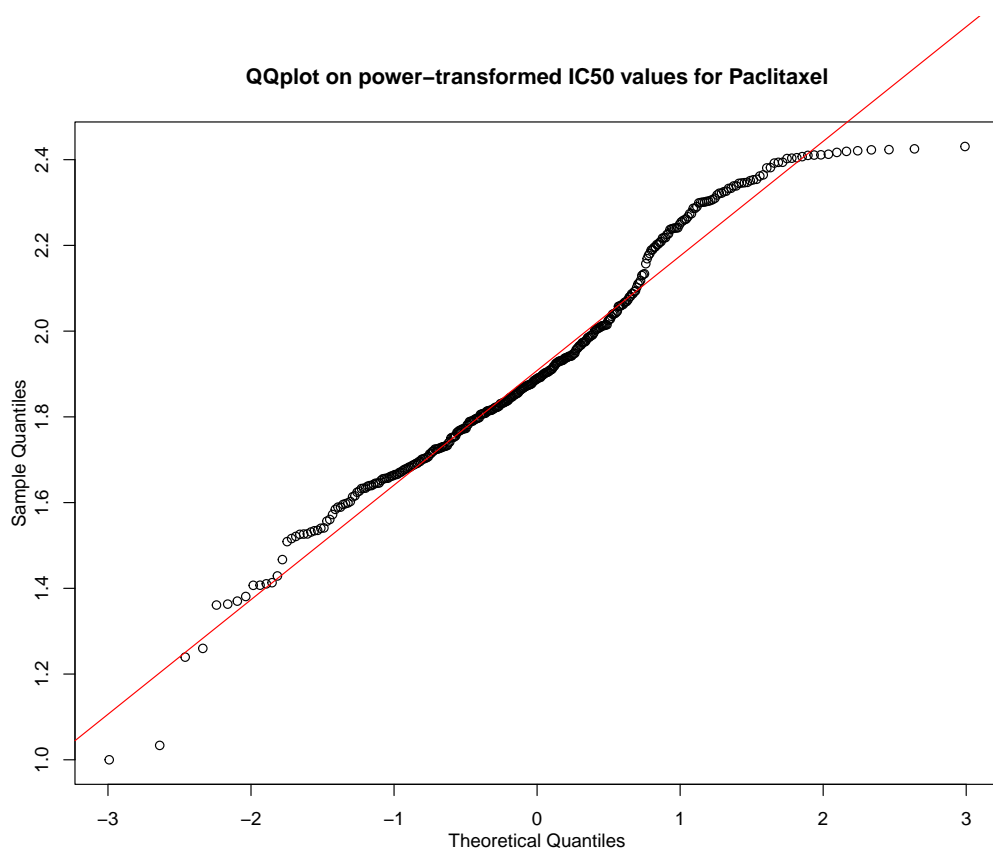


Figure 6: QQ plot for distribution of the transformed IC50 data

Figure 7 (下方图) 为图 BR estimate prediction accuracy 概览。

(对应文件为 [Figure+Table/BR-estimate-prediction-accuracy.pdf](#))

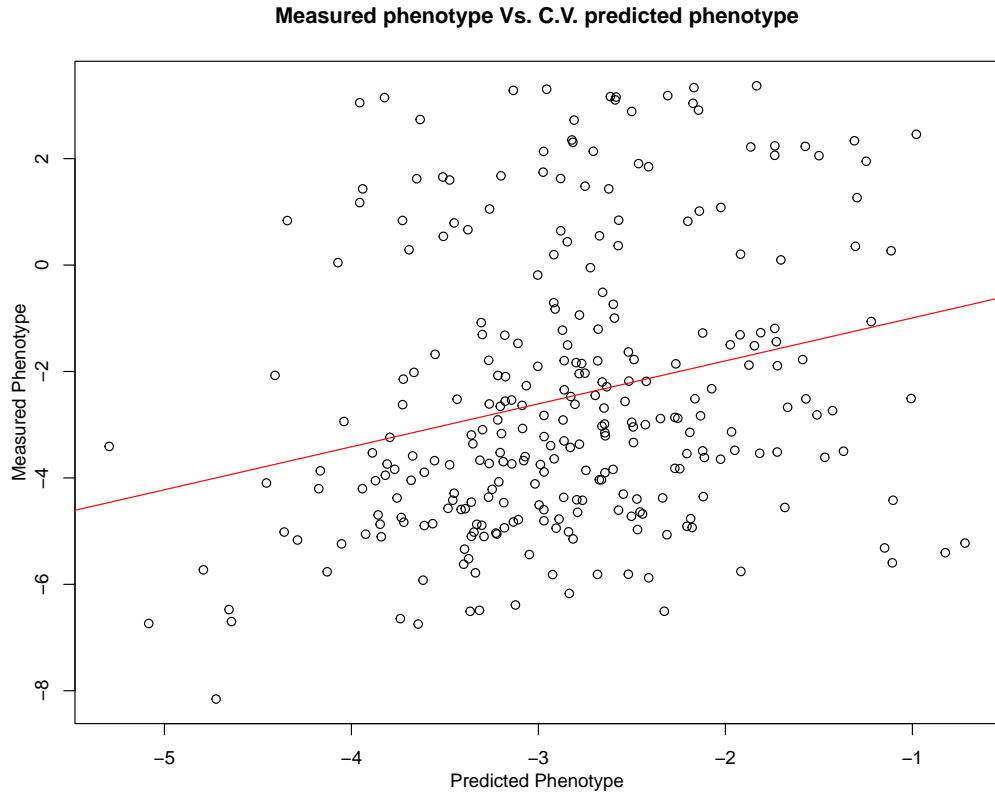


Figure 7: BR estimate prediction accuracy

Table 5 (下方表格) 为表格 BR predicted drug sensitivity 概览。

(对应文件为 Figure+Table/BR-predicted-drug-sensitivity.csv)

注：表格共有 229 行 3 列，以下预览的表格可能省略部分数据；含有 229 个唯一 ‘sample’。

1. sample: 样品名称

k-means clustering :

Centers = 3

Table 5: BR predicted drug sensitivity

sample	sensitivity	kmeans_group
TCGA-A1-A0SK-01A	-2.10842613469246	2
TCGA-A1-A0SO-01A	-2.09562910762259	2
TCGA-A1-A0SP-01A	-3.25995181471472	3

sample	sensitivity	kmeans_group
TCGA-A2-A04P-01A	-3.23445178824064	3
TCGA-A2-A04T-01A	-2.35183449260819	1
TCGA-A2-A04U-01A	-3.69008335272504	3
TCGA-A2-A0CM-01A	-3.52641697655325	3
TCGA-A2-A0D0-01A	-3.99448786955298	3
TCGA-A2-A0D2-01A	-3.5154928620097	3
TCGA-A2-A0EQ-01A	-2.44757901693141	1
TCGA-A2-A0ST-01A	-3.07631191508686	1
TCGA-A2-A0SX-01A	-3.29857199124933	3
TCGA-A2-A0T0-01A	-2.62869725258927	1
TCGA-A2-A0T2-01A	-2.86966880241339	1
TCGA-A2-A0YE-01A	-2.80843853558055	1
...

7.1.3 差异分析

成功筛选到 ABCB1, 见 Tab. 8

Table 6 (下方表格) 为表格 metadata 概览。

(对应文件为 **Figure+Table/metadata.xlsx**)

注：表格共有 229 行 98 列，以下预览的表格可能省略部分数据；含有 229 个唯一 ‘rownames’。

1. sample: 样品名称
2. group: 分组名称

Table 6: Metadata

rownames	group	lib.size	norm.f...	sample	barcode	patient	shortL...	defini...	sample...
TCGA-A...	Resist...	671449...	1.0768...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Resist...	701385...	1.0274...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	569427...	0.9859...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	442163...	0.8182...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Others	615423...	0.9621...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	384309...	0.9349...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	455529...	1.0497...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	437885...	0.8295...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	510092...	0.9319...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Others	507124...	0.9101...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...

rownames	group	lib.size	norm.f...	sample	barcode	patient	shortL...	defini...	sample...
TCGA-A...	Others	520602...	0.9950...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Non_re...	701598...	1.0795...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Others	581753...	0.9089...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Others	452328...	0.8451...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
TCGA-A...	Others	622011...	1.0001...	TCGA-A...	TCGA-A...	TCGA-A...	TP	Primar...	TCGA-A...
...

Figure 8 (下方图) 为图 BR Resistance vs Non resistance DEGs 概览。

(对应文件为 **Figure+Table/BR-Resistance-vs-Non-resistance-DEGs.pdf**)

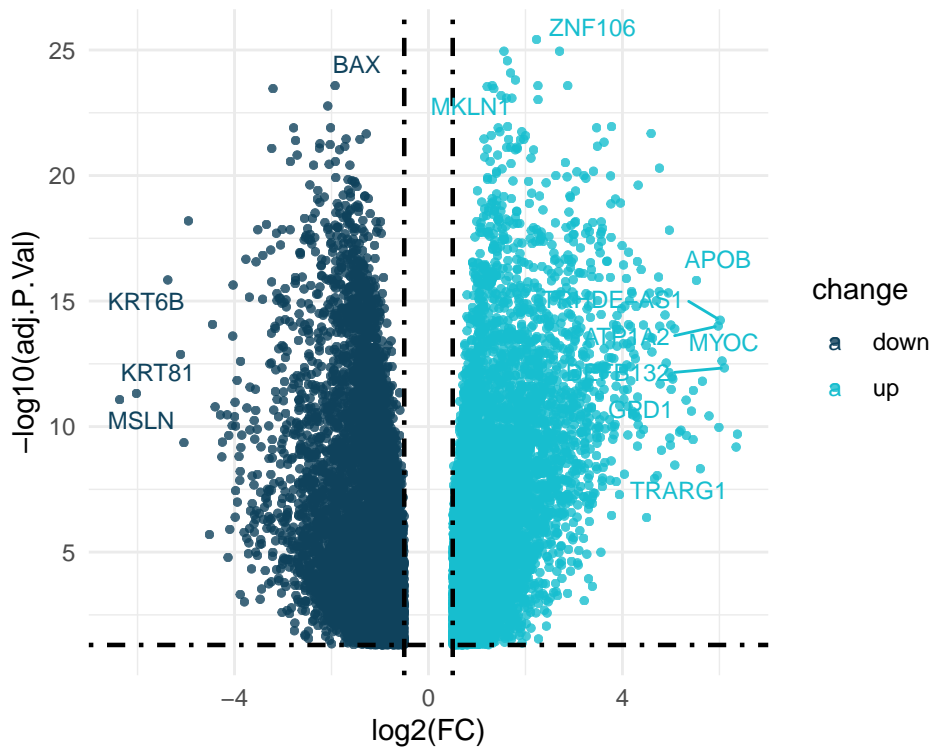


Figure 8: BR Resistance vs Non resistance DEGs

adj.P.Val cut-off :

0.05

Log2(FC) cut-off :

0.5

(上述信息框内容已保存至 **Figure+Table/BR-Resistance-vs-Non-resistance-DEGs-content**)

Table 7 (下方表格) 为表格 BR data Resistance vs Non resistance DEGs 概览。

(对应文件为 **Figure+Table/BR-data-Resistance-vs-Non-resistance-DEGs.csv**)

注：表格共有 13607 行 22 列，以下预览的表格可能省略部分数据；含有 13607 个唯一 ‘rownames’。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
2. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
3. t: moderated t-statistic (omitted for ‘topTableF’)
4. P.Value: raw p-value
5. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’)
6. gene_id: GENCODE/Ensembl gene ID
7. gene_name: GENCODE gene name
8. strand: genomic strand

Table 7: BR data Resistance vs Non resistance DEGs

rownames	gene_id	seqnames	start	end	width	strand	source	type	score
ENSG00...	ENSG00...	chr15	42412823	42491141	78319	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr4	84669597	84966690	297094	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr8	29055935	29056685	751	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr5	75511756	75601144	89389	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr17	46983287	47100323	117037	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr5	119037772	119249138	211367	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr7	131110096	131496632	386537	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr19	48954815	48961798	6984	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr10	116671192	116850251	179060	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr10	94402541	94536332	133792	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr11	392614	404908	12295	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr10	118004916	118046941	42026	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr4	107863473	107989679	126207	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr2	169827454	170084131	256678	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr11	9778667	10294219	515553	-	HAVANA	gene	NA
...

Table 8 (下方表格) 为表格 BR data Resistance vs Non resistance DEGs ABCB1 概览。

(对应文件为 **Figure+Table/BR-data-Resistance-vs-Non-resistance-DEGs-ABCB1.csv**)

注：表格共有 1 行 22 列，以下预览的表格可能省略部分数据；含有 1 个唯一 ‘rownames’。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
2. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
3. t: moderated t-statistic (omitted for ‘topTableF’)
4. P.Value: raw p-value
5. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’)
6. gene_id: GENCODE/Ensembl gene ID
7. gene_name: GENCODE gene name
8. strand: genomic strand

Table 8: BR data Resistance vs Non resistance DEGs ABCB1

rownames	gene_id	seqnames	start	end	width	strand	source	type	score
ENSG00...	ENSG00...	chr7	87503017	87713323	210307	-	HAVANA	gene	NA

Reference

1. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
2. Stelzer, G. *et al.* The genecards suite: From gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54**, 1.30.1–1.30.33 (2016).
3. Geeleher, P., Cox, N. & Huang, R. S. PRRophetic: An r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS one* **9**, (2014).
4. Szklarczyk, D. *et al.* The string database in 2021: Customizable proteinprotein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**, D605–D612 (2021).
5. Chin, C.-H. *et al.* CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**, S11 (2014).
6. Luo, W. & Brouwer, C. Pathview: An r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics (Oxford, England)* **29**, 1830–1831 (2013).