RNA Sequencing Uncovers Molecular Mechanisms Underlying Pathological Complete Response to Chemotherapy in Patients with Operable Breast Cancer

Yongfeng Li, Xiaozhen Liu, Hongchao Tang, Hongjian Yang, Xuli Meng

Background: This study aimed to identify key genes contributing to pathological complete response (pCR) to chemotherapy by mRNA sequencing (RNA-seq).

Material/Methods: RNA was extracted from the frozen biopsy tissue of patients with pathological complete response and patients with non-pathological complete response. Sequencing was performed on the HiSeq2000 platform. Differentially expressed genes (DEGs) were identified between the pCR group and non-pCR (NpCR) group. Pathway enrichment analysis of DEGs was performed. A protein-protein interaction network was constructed, then module analysis was performed to identify a subnetwork. Finally, transcription factors were predicted.

Results: A total of 673 DEGs were identified, including 419 upregulated ones and 254 downregulated ones. The PPI network constructed consisted of 276 proteins forming 471 PPI pairs, and a subnetwork containing 18 protein nodes was obtained. Pathway enrichment analysis revealed that PLCB4 and ADCY6 were enriched in pathways renin secretion, gastric acid secretion, gap junction, inflammatory mediator regulation of TRP channels, retrograde endocannabinoid signaling, melanogenesis, cGMP-PKG signaling pathway, calcium signaling pathway, chemokine signaling pathway, cAMP signaling pathway, and rap1 signaling pathway. CNR1 was enriched in the neuroactive ligand-receptor interaction pathway, retrograde endocannabinoid signaling pathway, and rap1 signaling pathway. The transcription factor-gene network consists of 15 transcription factors and 16 targeted genes, of which 5 were downregulated and 10 were upregulated.

Conclusions: We found key genes that may contribute to pCR to chemotherapy, such as PLCB4, ADCY6, and CNR1, as well as some transcription factors.

MeSH Keywords: Achaete-Scute Complex Genome Region • Breast Neoplasms, Male • Clinical Conference

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/903272
Background

Breast cancer is a leading gynecologic cancer. Currently, neoadjuvant chemotherapy (NCT) is widely used for the treatment of early breast cancer [1]. However, response to chemotherapy varies among patients, even when they have identical histological and molecular subtypes [2,3]. Pathological complete response (pCR) has been implicated as a surrogate indicator of long-term clinical benefit, such as disease-free survival and overall survival (OS), by many studies [4,5]. To date, only negative estrogen receptor (ER) status [6] and loss of bcl-2 [7] have been suggested as predictors of pCR in patients receiving neoadjuvant chemotherapy for operable and locally advanced breast cancer.

However, the mechanisms underlying pCR are largely unknown. Here, we attempted to use RNA sequencing (RNA-seq) technique to seek key genes and transcription factors that may contribute to pCR by comparing the transcriptome data between breast cancer patients with pCR to chemotherapy and those with non-pCR.

Material and Methods

Patient enrollment and sampling

Patients diagnosed with invasive breast cancer were included in this study. The inclusion criteria were: (1) patients with stage II or IIIA, operable breast cancer; (2) without distant metastasis; (3) without other serious disease, who can complete the chemotherapy as planned; (4) at least 10 ug total RNA can be extracted from the frozen biopsy tissue.

Finally, 6 patients, including 3 showing pCR (MXL-3-2, MXL-5-1 and MXL-P2) and 3 showing non-pCR (MXL-N1, MXL-N2 and MXL-1-3), were included. Written informed consent was provided by each patient and volunteer before sampling.

Total RNA extraction, library construction, and sequencing

A total of 10 g RNA was extracted from the frozen biopsy tissue of each patient, then poly(A)+ mRNA was isolated using Sera-mag Magnetic Oligo(dT) Beads (Thermo, part # 1004815). Afterwards, the purified RNA was broken into short segments in Fragment buffer. Next, reverse transcription was performed to construct a cDNA library. RNA concentration was measured with a Qubit® 2.0 Fluorometer and RNA integrity number (RIN) was measured by use of a Bioanalyzer 2100 (Agilent, CA, USA). Sequencing was performed on the HiSeq2000 platform (Illumina Inc., San Diego, CA) and single-end reads (50 bp in length) were generated.

Sequence quality control and alignment

First, reads were filtered by removing reads containing more than 40% bases with quality value <Q20 and reads containing more than 4% N bases. Sequence quality control was done with the fastx_toolkit. Next, clean reads were mapped to the human reference genome hg19 using TopHat 2.1.1 software (download site: http://ccb.jhu.edu/software/tophat/index.shtml).

Identification of differentially expressed genes correlated with mild and severe pneumonia

Reads count data were first preprocessed using the edgeR package of R (version 3.1.0) [8], which were then processed into log2-counts per million (logCPM) matrix by voom method using the R/limma package. Then, differential expression analysis between the pCR and NpCR groups was performed using the R/limma package, and | log2 FC | >1 and p<0.05 were used as cutoffs for a differentially expressed gene (DEG). A heat map was also used to display the gene expression profile of differentially expressed genes, based on the hierarchical clustering results using R package gplots [9].

Construction of a protein-protein interaction network

DEGs were submitted to the online tool STRING [10] (version: 10.0; http://string-db.org/) to identify protein-protein interaction (PPI) pairs. Only pairs with PPI score of 0.7 or higher were retained to construct a PPI network using Cytoscape [11].

Key proteins in the PPI network were identified by comparing the following parameters: degree of centrality [12], betweenness centrality [13], subgraph centrality [14], and closeness Centrality [15]. The parameters were calculated using a cytoscape plug-in CytoNCA [16].

Next, module analysis was performed to indent the subnetwork from the PPI network using ClusterONE [17]. Pathway enrichment analysis of the genes in the module was also performed using DAVID; p<0.05 was used as a cutoff.

Prediction of transcription factors and construction of a transcription factor-gene network

Transcriptional regulators are responsible for the transcriptional regulation of gene expression [18]. Transcriptional factors that may regulate the DEGs in a module were searched by use of the anplug-in iRegulon [19] of Cytoscape, which integrates transcription factor information from the databases Transfacs, Jaspar, Encode, Swissregulon, and Homer. The parameters were set as follows: Minimum identity between orthologous genes: 0.05, and Maximum false discovery rate on motif
similarity: 0.001. A gene-transcription factor pair with normalized enrichment score (NES) >3 was retained.

### Result

#### Quality control of reads and statistics

Quality statistics of reads are shown in Table 1.

#### Identification of DEGs

A total of 673 DEGs were identified in patients with pCR to chemotherapy, including 419 upregulated ones and 254 downregulated ones. Overall, the identified DEGs can distinguish the pneumonia samples from the control sample (Figure 1).

The upregulated DEGs were significantly enriched in 12 KEGG pathways, and the downregulated ones were significantly enriched in 2 pathways.

#### PPI network construction and module analysis

The constructed PPI network consisted of 276 proteins forming 471 PPI pairs (Figure 2). The top 10 proteins for each centrality parameter are listed in Table 2.

---

**Table 1.** Quality statistics of reads.

|                | MXL-1-3     | MXL-N1     | MXL-N2     | MXL-3-2     | MXL-5-1     | MXL-P2     |
|----------------|-------------|------------|------------|-------------|-------------|------------|
| Total reads    | 13118494    | 9277266    | 8370271    | 10284940    | 10998042    | 9655503    |
| Length reads   | 50          | 51         | 51         | 50          | 50          | 51         |
| Q20            | 13099904    | 9230891    | 8334022    | 10269116    | 10982423    | 9607845    |
| Q20%           | 99.86%      | 99.50%     | 99.57%     | 99.85%      | 99.86%      | 99.51%     |
| Q10            | 13118494    | 9276514    | 8369620    | 10284940    | 10998042    | 9654775    |
| Q10%           | 100.00%     | 99.99%     | 99.99%     | 100.00%     | 100.00%     | 99.99%     |
| Clean reads    | 10690546    | 6365594    | 5792665    | 8402125     | 8830451     | 6619228    |
| Valid ratio    | 81.49%      | 68.61%     | 69.21%     | 81.69%      | 80.29%      | 68.55%     |

**Figure 1.** Heat map showing the gene expression profile of differentially expressed genes based on their hierarchical clustering. The upregulated genes and downregulated genes are indicated by red and green, respectively.
One subnetwork was obtained, which included 18 protein nodes forming 79 protein-protein interaction pairs (Figure 3). More importantly, the heatmap revealed that these 18 genes can accurately distinguish the pCR samples and NpCR samples (Figure 4). Pathway enrichment analysis revealed that \textit{PLCB4} and \textit{ADCY6} were enriched in the renin secretion pathway, gastric acid secretion pathway, gap junction pathway, inflammatory mediator regulation of TRP channels pathway, retrograde endocannabinoid signaling pathway, melanogenesis pathway, cGMP-PKG signaling pathway, calcium signaling pathway, chemokine signaling pathway, cAMP signaling pathway, and rap1 signaling pathway. \textit{CNR1} was enriched in the

Table 2. The top 10 proteins for each centrality parameter in the protein-protein interaction network.

| Subgraph | Degree | Betweenness | Closeness |
|----------|--------|-------------|-----------|
| MCHR1   | 3613.0657 | 17 | MAPK14 7847.9087 | MAPK14 0.012748 |
| FPR2    | 3416.2122 | MCHR1 16 | PLCG1 6810.7944 | EDN1 0.012742 |
| EDN1    | 3354.0837 | ADCY6 15 | PPP2R1A 5778.3564 | PLCG1 0.012732 |
| AVP     | 2874.9324 | PLCG1 15 | MTOR 5671.6304 | PIK3CD 0.01272 |
| GGR     | 2680.8079 | PIK3CD 15 | MAP2K2 5196.929 | BDNF 0.012717 |
| GNA14   | 2441.132 | FPR2 14 | EDN1 5053.5815 | MAP2K2 0.012717 |
| CCK     | 2312.8115 | TRIO 14 | BDNF 4729.795 | GNA14 0.012713 |
| PLCB4   | 2118.2554 | AVP 13 | POLR2H 4465.244 | MTOR 0.012711 |
| TRIO    | 2074.557 | PLCB4 13 | ADCY6 4349.1655 | PXN 0.012711 |
| EDN3    | 2024.1368 | MAPK14 13 | CTNNA1 3852.7537 | AVP 0.012703 |
neuroactive ligand-receptor interaction pathway, the retrograde endocannabinoid signaling pathway, and the rap1 signaling pathway (Table 3).

Construction of transcription factor-gene network

The transcription factor-gene network consisted of 15 transcription factors and 16 targeted genes (Figure 5). These transcription factors were also differentially expressed: 5 downregulated and 10 upregulated.

Discussions

Because pCR to chemotherapy is recognized as an independent prognostic and predictive factor for survival of BC patients, we attempted to learn more about what contributes to the difference between patients with pCR and those with NpCR. Based on the transcriptome data from RNA-seq, we first identified DEGs that were associated pCR, and then tried to find key proteins (such as PLCB4, ADCY6, CNR1, and MAPK14) to from a protein-protein interaction perspective.

Among the upregulated genes, PLCB4 and ADCY6 were speculated to contribute to pCR to chemotherapy via multiple pathways, such as via renin secretion, gastric acid secretion, gap junction, inflammatory mediator regulation of TRP channels, retrograde endocannabinoid signaling, melanogenesis, cGMP-PKG signaling pathway, calcium signaling pathway, chemokine signaling pathway, cAMP signaling pathway, and rap1 signaling pathway. CNR1 was a downregulated gene, which was enriched in pathways of neuroactive ligand-receptor interaction, retrograde endocannabinoid signaling, and rap1 signaling pathway.

ADCY6 encodes adenyl cyclase 6, which is a membrane-associated enzyme catalyzing the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). The differential expression of this gene has also been reported in breast cancer [20] and ovarian carcinoma by microarray analysis [21]. By microarray analysis, ADCY6 expression was found to be decreased by 2-fold more in the 81-gemcitabine-resistant
variant MiaPaCa2-RG than in the gemcitabine-sensitive cell line MiaPaCa2 [22], suggesting that high ADCY6 expression is associated with gemcitabine sensitivity. This seems to agree with its upregulation in patients with pCR to chemotherapy.

PLCB4 encodes an enzyme 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-4, which catalyzes the formation of inositol 1,4,5-trisphosphate (IP3) and diacylglycerol from phosphatidylinositol 4,5-bisphosphate (PIP2). Woodfield et al. identified the primary gene targets of TFAP2C in hormone responsive breast carcinoma cells and found that PLCB4 was induced by TFAP2C [23]. TFAP2C is a member of the retinoic acid-inducible developmentally regulated family of AP-2 factors, which regulates its target genes via a helix-loop-helix motif [24] and has been suggested to have a critical role in breast cancer [25–27]. Thus, we supposed that PLCB4 may also have a role in the pathogenesis of breast cancer. Zheng et al. reported that PLCB4 was among the 12 genes showing significant upregulated expression (ratio >5) in acute myeloid leukemia cells (HL-60/MDR) with resistance to 6 ordinary antitumor drugs (Vincristine, Adriamycin, Etoposide, Daunorubicin, Mitomycin C, and Cyclophosphamide) [28], suggesting that PLCB4 contributes to resistance to antitumor drugs. However, this seems contradictory to our finding here that PLCB4 was upregulated in patients with complete response to chemotherapy. Thus, whether this helps to induce pCR to chemotherapy needs to be further studied. Further pathway enrichment analysis suggests that this gene may contribute to pCR to chemotherapy.

CNR1 was downregulated and encodes a G protein-coupled cannabinoid receptor (also known as cannabinoid CB1), located primarily in the central nervous system, and its activation is necessary for the antiemetic action of cannabinoids used for preventing emesis in cancer patients receiving chemotherapy [29]. Remarkably, decrease in its expression has also been reported in chemosensitive serous epithelial ovarian cancer tumors [30], suggesting that this gene may also have a role in breast cancer. Interestingly, these 3 genes were predicted to be regulated by transcription factors such as BARX2, CRX, TOX, POU2F1, VAX2, HOXA6, HOXB7, MYBL1, and NFATC2. Some of these transcription factors were upregulated and some were downregulated.

**Table 3.** KEGG pathway enrichment of genes in the protein-protein interaction subnetwork.

| Term                                           | P value   | Genes                      |
|------------------------------------------------|-----------|----------------------------|
| hsa04080: Neuroactive ligand-receptor interaction | 1.05E-07  | MCHR1, SSTR2, CNR1, FPR2, VIPR1, ADORA1, HTR2C, GCCR |
| hsa04924: Renin secretion                       | 5.28E-03  | PLCB4, ADCY6, ADORA1       |
| hsa04971: Gastric acid secretion                | 6.82E-03  | SSTR2, PLCB4, ADCY6        |
| hsa04540: Gap junction                          | 9.79E-03  | PLCB4, ADCY6, HTR2C        |
| hsa04750: Inflammatory mediator regulation of TRP channels | 1.18E-02  | PLCB4, ADCY6, HTR2C        |
| hsa04723: Retrograde endocannabinoid signaling  | 1.18E-02  | PLCB4, CNR1, ADCY6         |
| hsa04916: Melanogenesis                         | 1.25E-02  | PLCB4, EDN1, ADCY6         |
| hsa04022: cGMP-PKG signaling pathway            | 3.11E-02  | PLCB4, ADCY6, ADORA1       |
| hsa04020: Calcium signaling pathway             | 3.58E-02  | GNA14, PLCB4, HTR2C        |
| hsa04062: Chemokine signaling pathway           | 3.96E-02  | PLCB4, CCR5, ADCY6         |
| hsa04024: cAMP signaling pathway                | 4.44E-02  | SSTR2, ADCY6, ADORA1       |
| hsa04015: Rap1 signaling pathway                | 4.99E-02  | PLCB4, CNR1, ADCY6         |

**Figure 5.** Transcription factor-gene network for subnetwork module. The upregulated proteins and downregulated proteins are indicated by red and green, respectively. Triangles represent the transcription factor and ellipses represents the target protein.
indicating that their regulation upon the expression of a gene is complicated. Among them, BARX2 was found to be downregulated in patients showing pCR to chemotherapy. BARX2 is located within a minimal region at 11q25, which has been reported to promote breast cancer progression [31]. Sellar et al. reported that this gene can modulate cisplatin sensitivity in human epithelial ovarian cancer [32]. In contrast, HOXA6 and HOXB7 were upregulated in patients showing pCR to chemotherapy. These 2 transcription factors are encoded by the homeobox (HOX) gene family, and were found to have an important role in embryogenesis and differentiation of adult cells [33]. Some of the HOX members (Hoxa9, Hoxa10, and Hoxa11) have been suggested to contribute to epithelial ovarian cancer [34].

Conclusions

By RNA-seq, we found some key genes that may contribute to pCR to chemotherapy, such as PLCB4, ADCY6, CNR1, as well as some transcription factors. Their roles need to be further validated by experimental evidence.

Disclosure of conflict of interest

None.